Journal of Personality and Social Psychology
2017, Vol. 113, No. 3, 430–452

0022-3514/17/$12.00 http://dx.doi.org/10.1037/pspi0000097

It Doesn't Hurt to Ask: Question-Asking Increases Liking

Karen Huang, Michael Yeomans, Alison Wood Brooks, Julia Minson, and Francesca Gino
Harvard University

Conversation is a fundamental human experience that is necessary to pursue intrapersonal and interpersonal goals across myriad contexts, relationships, and modes of communication. In the current research, we isolate the role of an understudied conversational behavior: question-asking. Across 3 studies of live dyadic conversations, we identify a robust and consistent relationship between question-asking and liking: people who ask more questions, particularly follow-up questions, are better liked by their conversation partners. When people are instructed to ask more questions, they are perceived as higher in responsiveness, an interpersonal construct that captures listening, understanding, validation, and care. We measure responsiveness with an attitudinal measure from previous research as well as a novel behavioral measure: the number of follow-up questions one asks. In both cases, responsiveness explains the effect of question-asking on liking. In addition to analyzing live get-to-know-you conversations online, we also studied face-to-face speed-dating conversations. We trained a natural language processing algorithm as a "follow-up question detector" that we applied to our speed-dating data (and can be applied to any text data to more deeply understand question-asking dynamics). The follow-up question rate established by the algorithm showed that speed daters who ask more follow-up questions during their dates are more likely to elicit agreement for second dates from their partners, a behavioral indicator of liking. We also find that, despite the persistent and beneficial effects of asking questions, people do not anticipate that question-asking increases interpersonal liking.

Keywords: question-asking, liking, responsiveness, conversation, natural language processing

Imagine this scenario: you meet a new colleague for the first time at a company party. You strike up a conversation, and the colleague starts telling you a funny story. You are interested and engaged, and you ask several questions that encourage the colleague to elaborate on the details of the story. After the story is over, you exchange pleasantries and part ways. Later you realize that your colleague didn't ask any questions about you, and you didn't have an opportunity to reveal much information about yourself. Who made the better impression?

This article was published Online First April 27, 2017.

Karen Huang, Harvard Business School and Department of Psychology, Harvard University; Michael Yeomans, Institute for Quantitative Social Science, Harvard University; Alison Wood Brooks, Harvard Business School, Harvard University; Julia Minson, Harvard Kennedy School, Harvard University; Francesca Gino, Harvard Business School, Harvard University.

We are grateful for insightful feedback from Eli Finkel, Patrick Mair, Dennis Zhang, Steve Worthington, Barbara Wood, the Harvard Business School's NERD Lab members, the University of Chicago's Behavioral Science seminar participants, and the Wharton School's Operations, Information and Decisions seminar participants; for data generously shared by Dan Jurafsky, Rajesh Ranganath, and Dan McFarland; and for research assistance from Ethan Ludwin-Peery, Ashley Kirsner, Jean Sohn, Holly Howe, Mackenzie Lowry, Elise Lee, Isabelle Moore, Ben Schenck, Logan Berg, and Yilin Chen.

For each study, we report how we determined our sample size, all data exclusions, all manipulations, and all measures. The exact data from each study are available as online supplemental materials at https://osf.io/8k7rf/

Correspondence concerning this article should be addressed to Karen Huang, Harvard Business School, Harvard University, Wyss House, Boston, MA 02163. E-mail: karenhuang@g.harvard.edu

Conversation is a pervasive human experience. Conversing with others is a fundamental behavior across myriad contexts, relationships, and modes of communication (e.g., written, spoken). People can choose from many ways to contribute to a conversation, including making a statement, telling a story, making a quip or joke, apologizing, giving a compliment, or saying nothing at all while a conversation partner speaks (Clark & Schaefer, 1989). We converse with others to learn what they know—their information, stories, preferences, ideas, thoughts, and feelings—as well as to share what we know while managing others' perceptions of us. That is, two central goals of conversation are information exchange and impression management. In this article, we examine an understudied conversational behavior that likely influences both of these goals: question-asking.

Although question-asking is ubiquitous, we know very little about the antecedents and consequences of asking questions during interpersonal interaction. In the current research, we investigate the psychology of question-asking as a social phenomenon. We measure people's natural rates of question-asking and explore how the propensity to ask questions influences interpersonal liking across controlled experimental settings and an observational field setting. Compared with people who ask few questions, we expect that high question askers are better liked. In particular, asking questions that follow up on the other person's responses may cause and convey better listening, understanding, validation, and care (i.e., responsiveness, Reis, Maniaci, Caprariello, Eastwick, & Finkel, 2011; Reis & Patrick, 1996). The question asker's responsiveness, in turn, is likely to cause him or her to be better liked by the question answerer.

Question-Asking in Conversation

A conversation is a cooperative interaction in which each person acts in coordination to contribute to a successful experience of shared understanding (Clark & Schaefer, 1989). It is an ongoing, sequential unfolding of actions and responses (Reis & Patrick, 1996), organized as speaker turns (Schegloff & Sacks, 1973). Most conversations are characterized by the transfer of information about beliefs, thoughts, or emotions from one person to another (Epley & Waytz, 2010). In the current work, we investigate the social phenomenon of asking questions that encourage the partner to elaborate on their beliefs, thoughts, and emotions.

Question-asking directs conversations by encouraging another person to answer (Dillon, 1982, 1988). Though some people may ask questions to avoid disclosing information themselves, most questions function to solicit information from others (Chafe, 1970; Dillon, 1982; Kearsley, 1976). If one person asks a question, the other person's response should abide by basic conversational maxims (Graesser, 1985; Grice, 1975), such as responding with the relevant information to the question at hand (Hilton, 1990). Although, some recent work suggests that people could violate these norms by dodging questions, responding with truth that is deliberately misleading (i.e., paltering), or refusing to answer altogether (John, Barasz, & Norton, 2016; Rogers & Norton, 2011; Rogers, Zeckhauser, Gino, Norton, & Schweitzer, 2016).

The type of question-asking we investigate—natural, conversational questions that elaborate on the question-responder's statements—differ categorically from the questions investigated in studies on experimentally induced social closeness (e.g., Aron, Aron, Tudor, & Nelson, 1991; Aron, Melinat, Aron, Vallone, & Bator, 1997; Sedikides, Campbell, Reader, & Elliot, 1999; Sprecher, Treger, Wondra, Hilaire, & Wallpe, 2013). This prior work has defined social closeness as the inclusion of the other in the concept of the self (Aron, Aron, & Smollan, 1992; Aron et al., 1991). In this work, participants were instructed to ask a fixed list of questions that change topic but increase in intimacy over time, and partners take turns answering all questions (e.g., Aron et al., 1997). For example, each partner would take turns asking and answering the question "What do you value most in a friendship?" before moving on to asking and answering the question "What is your most treasured memory?" (Aron et al., 1997). In these studies, questions were provided by an experimenter, and participants were not instructed or encouraged to ask follow-up questions. In contrast, in our work, we investigate the effect of question-asking on liking in natural dyadic interactions.

We focus on information-seeking questions (e.g., Miles, 2013; Van der Meij, 1987) in which the question-asker lacks some information and requests more information from the other person. People often ask information-seeking questions when meeting for the first time (Berger & Calabrese, 1975), and are more likely to seek information from others when they consider the information highly valuable (Swann, Stephenson, & Pittman, 1981). Because people often know very little about each other upon first meeting, individuals stand to learn a large amount of information about their conversation partners during first encounters. Importantly, though, information exchange is not the only goal of conversation. Asking questions may serve and influence other motivations like impression management.

Question-Asking and Liking

Most people have an intrinsic desire to be liked by others (Baumeister, 1982; Jones & Pittman, 1982; Leary & Kowalski, 1990). Being liked by others influences interpersonal attraction, relationship development (Berscheid, 1985; Berscheid & Regan, 2005), and other important outcomes such as acceptance and inclusion in groups (Reis & Patrick, 1996).

Because the content of a conversation can significantly influence the extent to which the participants like each other afterwards, it is important to examine conversation as a process that influences attraction (Davis & Perkowitz, 1979) and relationship development (Reis & Shaver, 1988; Miller, Berg, & Archer, 1983). The effect of conversational content on interpersonal liking has been demonstrated across a wide array of conversational strategies, ranging from other-focused behaviors, such as giving a compliment or acknowledging another person's ideas, to self-focused behaviors, such as talking about oneself (Godfrey, Jones, & Lord, 1986; Laurenceau, Barrett, & Pietromonaco, 1998; Rosenfeld, 1966; Sprecher et al., 2013). However, to our best knowledge, no prior research has investigated whether and how asking questions may influence liking.

Though asking questions invites information disclosure, there are many reasons why people may not ask questions. First, people may not think to ask questions at all. Neglecting to ask questions altogether may happen because people are egocentric—focused on expressing their own thoughts, feelings, and beliefs (e.g., Gilovich, Medvec, & Savitsky, 2000) with little or no interest in hearing what another person has to say. Or they may be too distracted by other aspects of the conversation (e.g., emotion expression) that they do not realize that asking a question is an option. On the other hand, some people may think to ask questions, but may purposefully forgo asking because they are unsure about which question(s) to ask or worry about asking a question that is perceived as rude, inappropriate, intrusive, or incompetent. In these cases, it may be much easier to talk about oneself instead.

Indeed, in most conversations, people predominantly share information about themselves rather than discussing other possible topics (Landis & Burtt, 1924). A study of conversations in public settings such as bars and trains suggests that people spend two thirds of conversation time talking about their personal experiences (Dunbar, Marriott, & Duncan, 1997). Especially when meeting someone new, people tend to use self-focused presentation strategies like self-promotion (Godfrey et al., 1986). For example, Marr and Cable (2014) found that job candidates excessively attempt to "sell" themselves to make a favorable impression in job interviews.

The tendency to focus on the self when trying to impress others is misguided, as verbal behaviors that focus on the self, such as redirecting the topic of conversation to oneself, bragging, boasting, or dominating the conversation, tend to decrease liking (Berman, Levine, Barasch, & Small, 2015; Godfrey et al., 1986; Sezer, Gino, & Norton, 2015; Vangelisti, Knapp, & Daly, 1990). In contrast, verbal behaviors that focus on the other person, such as mirroring the other person's mannerisms (Ireland & Pennebaker, 2010), affirming the other's statements, or coaxing information from the other person, have been shown to increase liking (Godfrey et al., 1986; Rosenfeld, 1966).

We hypothesize that asking more questions—and in particular, asking more follow-up questions—increases liking for the question asker. This hypothesis is consistent with prior research. For example, at the trait level, people who tend to draw out more information from their conversation partners (termed "openers") are better liked by their partners in long-term relationships (Miller et al., 1983). And studies of doctor–patient communication suggest that patients report higher satisfaction with their visits when physicians ask more questions about the patients' experiences (Bertakis, Roter, & Putnam, 1991; Robinson & Heritage, 2006). Furthermore, because most people spend the majority of their conversations sharing their own views rather than focusing on the other person, we hypothesize that people do not anticipate the effect of question-asking on liking.

Responsiveness Mediates the Effect of Question-Asking on Liking

We suggest that asking questions increases liking because doing so indicates *responsiveness*, a desirable interpersonal construct identified by prior research that encompasses the verbal and nonverbal behaviors that fulfill the needs and wishes of one's conversation partner (Davis, 1982; Miller & Berg, 1984). Responsive behavior in a conversation requires a set of skills for responding relevantly and appropriately. We argue that question-asking is one conversational behavior that is likely to convey high responsiveness.

Reis and Shaver (1988) developed a model of interpersonal intimacy that defines responsiveness as reflecting three components: understanding, validation, and care for the partner. First, the *understanding* component of responsiveness refers to accurately comprehending the question-responder's self-perceptions—their needs, goals, beliefs, emotions, and life situation (Reis & Patrick, 1996; Reis & Shaver, 1988). By asking questions, one elicits information from the partner, including facts, attitudes, preferences, and emotional expressions, which help to more accurately and appropriately understand one's partner. Understanding cannot take place without being well-informed about one's partner (Reis & Patrick, 1996), and question-asking is likely to increase the disclosure and learning necessary for understanding.

Second, the *validation* component of responsiveness is defined as valuing and respecting the partner's self-perceptions and perspectives (Reis & Patrick, 1996; Reis & Shaver, 1988). Validation also involves affirming that the partner is accepted and valued (Reis & Shaver, 1988). We suggest that asking questions communicates respect and value for the partner's perspective. Ironically, even without responding to the partner with direct validation or affirmation, question-asking itself may be seen as a form of positive approval or validation (Cozby, 1973). By asking questions, you acknowledge that the partner's perspective is valuable enough that you want to know more. By soliciting more information from the partner, asking a question expresses interest in the partner's viewpoint (Chen, Minson, & Tormala, 2010). Indeed, previous research suggests that effective validation in marital communication can be successfully conveyed by asking open-ended questions (Notarius & Markman, 1981).

Finally, the *caring* component of responsiveness means showing affection and concern for the partner (Reis & Patrick, 1996; Reis & Shaver, 1988). Especially in initial interactions that are often devoid of prior relational information, asking questions is likely to signal care for the partner. Rather than talking about oneself, asking questions about the partner is likely to indicate warmth, positive affect, curiosity, and empathic concern—the question asker shows that he cares to know about the conversation partner's perspective. Expressing affection and care for the partner tends to increase liking by the partner, due to reciprocity (Montoya & Insko, 2008; Gouldner, 1960; Wilson & Henzlik, 1986; Sprecher, 1998).

According to Reis and Patrick's (1996) model of responsiveness, understanding is often a necessary requirement of validation and care. That is, one cannot validate and care for someone without first accurately recognizing and acknowledging his or her self-perceptions. In a study that manipulated understanding and validation orthogonally, liking increased for validating partners when they were accurate rather than inaccurate (Patrick & Reis, 1995; Reis & Patrick, 1996). One needs to first accurately understand their partner's beliefs and attitudes in order to validate them.

The construct of responsiveness aligns closely with the concept of active listening discussed in fields such as communication and marital therapy (e.g., Bodie, 2011; Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012; Gordon, 1975; Lester, 2002; Rogers, 1951; Stanley, Bradbury, & Markman, 2000; Weger, Bell, Minei, & Robinson, 2014). Like the understanding and validation components of responsiveness, active listening requires paying full attention to the partner in the conversation (Bodie, 2011; Hutchby, 2005; Rogers, 1955; Rogers & Farson, 2007), and most definitions and studies of active listening emphasize the importance of asking questions that are relevant to the partner's statements (Paukert, Stagner, & Hope, 2004; Weger et al., 2014; Bodie, 2011; Minkin et al., 1976). A listener's responses regulate the conversation (Duncan & Fiske, 1977; Patterson, 1994), such that responsive verbal behaviors can improve the fluency of the conversation, while unresponsive behaviors can end the conversation (Davis, 1982).

Taken together, we expect perceptions of responsiveness—understanding, validation, and care—to mediate the relationship between question-asking and liking. Asking more questions is likely to increase perceptions of responsiveness, and perceptions of responsiveness, in turn, are likely to increase interpersonal liking. Consistent with this theoretical model, the effect of question-asking on liking may only hold when people ask more follow-up questions, rather than other types of questions.

We define follow-up questions as questions that encourage the partner to elaborate on the content of their prior conversational turn (Davis, 1982). This definition underscores previous conceptualizations of follow-up questions identified in the active listening literature (Paukert et al., 2004; Weger et al., 2014). Follow-up questions are only possible if an individual asks an original question, listens to the answer, and probes for more information (i.e., understands the answer, validates the partner, and cares to know more—the definition of responsiveness). Thus, we predict that one's follow-up question rate is associated with higher liking from the question-answerer toward the question-asker.

Further, the effect of question-asking on liking may only influence liking of the question-asker by the conversation partner himself (Davis & Perkowitz, 1979). Because we expect the benefits of question-asking on liking to be explained by responsiveness to the conversation partner, we predict that increased question-asking will not influence liking by third-party observers of the conversation.

Overview of the Current Research

In a series of four studies, we investigate the patterns and effects of question-asking in dyadic conversation. In Study 1, we instruct one conversation partner in a dyad to ask a high or low number of questions and measure the other partner's liking of the question asker. In Study 2A, we manipulate high or low question-asking for both conversation partners. In both Study 1 and 2A, we investigate responsiveness as a psychological mechanism underlying the main effect. In Study 2B, we ask third-party observers to rate conversation partners on liking. Furthermore, we conduct a joint analysis of the types of questions people asked in Studies 1 and 2 to investigate the effect of follow-up questions on liking. Finally, in Study 3, we investigate the effect of question-asking in a field context (speed-dating) with a behavioral measure of liking (being asked on a second date), and we develop a natural language processing algorithm that can classify question types automatically in any conversation data.

Analytical Strategy

The studies in this article span a wide range of designs and methods. In general, we conducted our analyses to test effects at the dyadic level—that is, how Person A's level of question-asking affects Person B's evaluations of person A, or how Person A's question-asking affects how Person A thinks they will be evaluated by Person B.

In Study 1, only one person in each dyad received a question-asking manipulation, and the other partner received no manipulation. We measured our outcomes of interest only once per dyad—that is, we measured the question-receiver's evaluation of the question-asker. Our analytic approach in Study 1 reflects this study design. In Study 2, both individuals in each dyad received the manipulation. Therefore, we measured outcomes of interest twice per dyad. Thus, we used mixed effects regression models, implemented though the *lme4* package in R (Bates, Maechler, Bolker, & Walker, 2014), to control for dyad-level variation.

In Study 3, we did not manipulate question-asking. Rather, we observed how individuals naturally asked and responded to questions, as they were paired on speed-dates with several other individuals. Therefore, we measured outcomes twice per dyad among people who participated in many dyads. The rate of question-asking may be correlated across a given individual's speed dates, especially if question-asking behavior is stable or trait-like. This correlation across dates required us to adjust all standard errors from our regressions to be robust in two ways: clustering within raters, and clustering within askers (Cameron, Gelbach, & Miller, 2011). We conducted this adjustment using the *multiwayvcov* package in R (Graham, Arai, & Hagstromer, 2016). Additionally, some models also include fixed effects—for askers, for raters, and for gender—to control for different sources of variation that affect partner liking in this domain.

For each study, we report how we determined our sample size, all data exclusions, all manipulations, and all measures (Simmons, Nelson, & Simonsohn, 2011). In our online supplemental materials on Open Science Framework, we provide the data and analysis code from each study.

Study 1

In Study 1, we test the effect of question-asking by randomly assigning one participant in a two-person conversation to ask either a high or a low number of questions. The other conversation partner did not receive or know about the question-asking manipulation. After the conversation, both participants reported how they felt about the conversation and their partner, and how they thought their partner felt about them. To investigate the psychological mechanism underpinning the relationship between question-asking and liking, we coded the responsiveness of each conversation partner.

Method

Participants. We recruited 430 participants (215 dyads) to participate in a "Chat Study" in a behavioral lab. We applied several exclusion criteria that were determined a priori to ensure our analysis only considered dyads in which both participants completed the full survey. Accordingly, we excluded three dyads in which at least one partner did not finish the study, three dyads in which at least one partner indicated that he or she was not paired with another participant, and 10 dyads in which at least one partner reported that he or she was not able to complete a full conversation. These exclusions left a sample of 398 participants (194 male, 204 female), or 199 dyads, for our analyses.

We recruited participants in three waves because of lab recruiting constraints. In one recruitment wave, participants completed only our study and were paid $15. In two other recruiting waves, participants completed our study among a bundle of unrelated studies. In these latter cases, participants were paid $20 and $25, respectively. We found no differences in our results controlling for recruitment wave and report our results collapsed across all three waves.

Design and procedure. We asked participants to sit in separate cubicles in our behavioral lab. All study materials were presented on computers that were separated by dividers, and participants did not interact face to face before, during, or after the experimental session. Instead, participants interacted by sending each other instant messages using an interface called ChatPlat, an application that enables experimenters to pair people easily and allow them to chat with each other within an online survey. ChatPlat has been used and validated in previous research (e.g., Brooks & Schweitzer, 2011).

Participants were paired with another participant in the room based on their arrival time at the ChatPlat chat window (i.e., the first-arriving participant was matched with second arriver, and so on). Participants were anonymous and unknown to one another. To get the conversation started, they both read these instructions: "You will chat for 15 minutes. During the conversation, your objectives are for you and your partner to get to know each other and learn about each other's interests." Participants were also told to pay attention during the chat because they would be asked to complete several questionnaires about their partner after they finished chatting. After chatting for 15 min, the chat window closed automatically. Participants received a notification one minute before the end of the chat.

After participants were paired, each dyad was randomly assigned to one of two conditions: *many-questions* or *few-questions*. In the many-questions condition, one participant in each dyad was told that he or she needed to ask "at least nine questions." In the few-questions condition, the participant was told that he or she needed to ask "at most four questions." These question-asking values were determined based on the natural base rate of question-asking from a separate pilot study conducted in the same behavioral laboratory ($N = 193$). We used the 25th (four questions) and 75th percentiles (nine questions) of question-asking from conversations in the pilot study to ensure the number of questions would be noticeably different from an average conversation ($M = 6.72$, $SD = 4.16$), but still natural. The participants who received question-asking instructions were also told not to let their partners know they had been given additional instructions. None of the participants were told the purpose of these instructions, and they were blind to our hypotheses.

Dependent variables. At the end of the chat, participants in all conditions reported their liking for their partner, and predicted their partner's liking of them, using the same four-item survey measure of interpersonal liking (see Appendix A for a full list of measures). In addition to liking, we also measured learning. We measured participants' knowledge of their partner using the Activity Preferences Questionnaire (APQ; Surra & Longstreth, 1990; Swann & Gill, 1997), a nine-item block of Likert responses that ask participants to indicate enjoyment of common activities (i.e., cooking, sports, reading, etc.). Each participant gave their own answers to the APQ and predicted how their partner would answer the APQ items (order counterbalanced). At the end of the survey, we included a manipulation check, asking participants if they were instructed to ask questions and, if so, how many.

Coding of conversations. We coded the text written by each participant for responsiveness. Coding of all 199 conversations was split among six research assistants, who were blind to condition and hypotheses, such that every conversation was coded by three independent raters. Research assistants read conversations in randomized order, and rated the degree to which they thought that each person in the conversation perceived their partner as responsive, on a 1 (*not at all*) to 7 (*very much so*) scale, using the perceived responsiveness scale described by Reis et al. (2011), which captures the three components of responsiveness: understanding, validation, and care.

Results

Question-asking. Throughout this article, we measure question-asking using a simple algorithm that counted conversational turns that included question marks. This method produced virtually identical results when compared with human coders (Cronbach's α: Study 1 = .95, Study 2 = .97). Using this scheme, if someone asked multiple questions in a single turn (i.e., before their partner responded), this was counted as a single question. However, the following results are identical if we account for multiple-question turns, as well. We used this algorithm to compute the total number of turns in which a question was asked (number of questions asked) as well as proportion of all conversational turns that included a question (question rate).

Consistent with our intended manipulation, participants who were instructed to ask many questions did in fact ask more questions ($M = 10.23$, $SD = 4.94$) than participants who were instructed to ask few questions ($M = 4.34$, $SD = 2.16$), two-sample t test: $t(197) = 10.87$, $p < .001$, Cohen's $d = 1.22$. Participants who received no instructions fell in between ($M = 7.03$, $SD = 3.95$). The same pattern held when questions were measured as a percentage of all conversational turns: Those assigned to ask many questions had a higher question rate ($M = 39.06\%$, $SD = 18.94\%$) than did those assigned to ask few questions ($M = 21.83\%$, $SD = 14.75\%$), two-sample t test: $t(197) = 7.16$, $p < .001$, Cohen's $d = .91$. Participants who received no instructions had a question rate that fell in between ($M = 27.75\%$, $SD = 14.46\%$). These results show that our question-asking instructions successfully manipulated high and low question-asking.

Liking. The primary dependent measure for this study was a block of four items about how much participants liked their partner after the conversation had ended (see Appendix A). These items were aggregated into a single standardized index of liking (Cronbach's $\alpha = .87$), and we plot the averages by condition in Figure 1.

We test our primary hypothesis by testing the effects of the high (vs. low) question-asking instructions on the partner who did not receive instructions. Because our sample consisted of 199 dyads, we used an independent sample t test to compare the average partner liking scores reported by the 199 dyad members who did not receive the manipulation, but instead interacted with partners who asked them a high or low number of questions. Confirming our prediction, participants paired with high question-askers liked their partners more ($M = 5.79$, $SD = 1.21$) than did participants paired with low question-askers ($M = 5.31$, $SD = 1.48$), $t(197) = 2.47$, $p = .014$; Cohen's $d = .35$. Not surprisingly, there was no difference in liking among those who received the instructions because the unmanipulated partners asked a similar number of questions in both conditions.[1] Those who were instructed to ask many questions liked their partners just as much ($M = 5.76$, $SD = .94$) as did participants who were instructed to ask few questions ($M = 5.67$, $SD = 1.27$), two-sample t test: $t(197) = .51$, $p = .612$.

Predicted liking. Predicted liking was reported on the same four items used to measure liking, but participants were asked to anticipate their partner's liking of them (Cronbach's $\alpha = .85$; see Appendix A). There was no difference in predicted liking between participants who were instructed to ask many questions ($M = 5.27$, $SD = .93$) or few questions ($M = 5.19$, $SD = 1.05$), $t(197) = .61$, $p = .544$). This (null) result suggests that individuals do not anticipate that a higher rate of question asking will lead to an increase in liking.

Our experimental design allows us to also answer the question of how the effect of question asking on the actual liking experienced by the unmanipulated partners compares to the predicted liking reported by the manipulated partners. In order to answer this question we test the 2 (manipulation: high vs. low question-asking) × 2 (perspective: unmanipulated partners' actual liking vs. manipulated partners' predicted liking) interaction, in a hierarchical linear model that controls for the fact that ratings were nested within each dyad. The interaction term suggested that the question-

[1] Examining the correlation between the question-asking rate of the unmanipulated participants and the liking reported by their partners, reveals a suggestive, but not significant, correlation ($r = .11$), $t(197) = 1.54$, $p = .125$. We return to this question in Study 3 with a larger data set of natural question-asking rates.

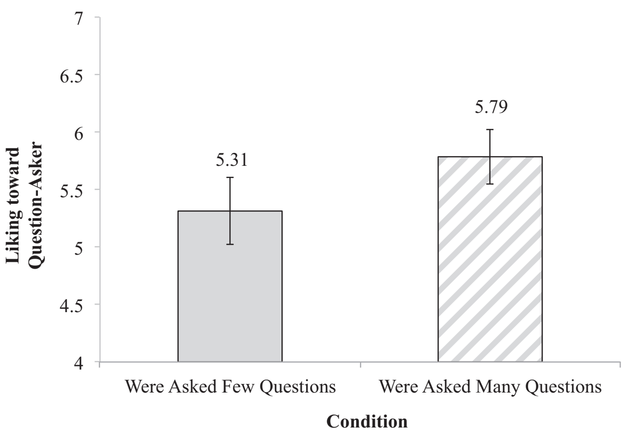

Figure 1. The effect of question-asking on liking in Study 1. In each pair, one person was randomly assigned to receive either few or many questions from the question-asker. Error bars represent 95% CI for the group means.

asking manipulation did indeed influence the partners' actual liking (marginally) more than the askers' prediction of that liking ($\beta = -.39$, $SE = .22$), $t(197) = 1.73$, $p = .084$ (see full model in Table 1, Panel A). Furthermore, across all conditions, there was no correlation between participants' predicted liking and their question-asking rate ($r = -.02$), $t(396) = .35$, $p = .724$. These results suggest that participants did not think that question-asking had an effect on liking.

Learning. We calculated the intraclass correlation (ICC) of the participants' predicted ratings and their partners' actual ratings on the nine APQ items, based on previous research (Shrout & Fleiss, 1979; Swann & Gill, 1997). The results remain unchanged if we use alternative metrics (e.g., difference scores, rank-order correlation). Though participants assigned to ask many questions were not significantly more accurate ($M = .33$, $SD = .32$) than were participants assigned to ask few questions ($M = .27$, $SD = .32$), two-sample t test: $t(197) = 1.26$, $p = .211$, there was a significant correlation between question-asking rate and learning ($r = .25$), $t(196) = 3.68$, $p < .001$, among those who did not receive question-asking instructions. It may be the case that the instructions to generate additional questions interfered with participants' ability to retain the information that their partners shared.

Responsiveness. There was high agreement among the coders on ratings of responsiveness (ICC = .75). In line with our hypotheses, participants who were instructed to ask many questions were rated as more responsive to their partner ($M = 4.68$, $SD = 1.08$) than participants who were instructed to ask few questions ($M = 4.37$, $SD = .99$), $t = 2.14$, $p = .034$; Cohen's $d = .30$. There was no difference in the rated responsiveness between the unmanipulated participants who were partnered with a high-question-asker ($M = 4.55$, $SD = .99$) and those who were partnered with a low-question-asker ($M = 4.61$, $SD = .98$), $t(197) = .38$, $p = .707$.

As a test of our proposed model, we conducted a mediation analysis using a nonparametric bootstrap sampling procedure (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). We estimated the causal pathway linking question-asking with liking for the ques-

Table 1
Full Hierarchical Linear Models (Studies 1 and 2)

Predictor variable	Estimate	*SE*	*t*	*p*
Panel A: Study 1 asker's predicted liking				
Condition (manipulated high vs. low question-asking)	.28	.13	2.22	.028
Perspective (partner's liking vs. asker's predicted liking)	−.32	.11	2.88	.004
Interaction term	−.39	.22	1.73	.084
Panel B: Study 2A asker's predicted liking				
Condition (manipulated high vs. low question-asking)	.11	.07	1.59	.113
Perspective (partner's liking vs. asker's predicted liking)	−.70	.06	12.40	<.001
Interaction term	−.29	.11	2.55	.011
Panel C: Study 2B observer liking				
Condition (manipulated high vs. low question-asking)	.14	.05	2.62	.009
Perspective (partner's liking vs. observer's liking)	−5.89	.03	213.08	<.001
Interaction term	−.18	.06	3.25	.001
Panel D: Study 2B observer liking				
Condition (manipulated high vs. low questions received)	−.08	.05	1.51	.132
Perspective (partner's liking vs. observer's liking)	−5.89	.03	213.43	<.001
Interaction term	.22	.06	4.02	<.001
Panel E: Study 2B observer predicted liking				
Condition (manipulated high vs. low question-asking)	.15	.05	2.93	.004
Perspective (partner's predicted liking vs. observer's predicted liking)	−5.89	.03	222.84	<.001
Interaction term	−.21	.05	3.95	<.001

Note. In the text we report the interaction terms from these models, which test whether a secondary measure of interest (e.g., predicted liking or third-party observer liking) tracks the effect of question-asking on the primary measure of interest (partner liking). For completeness we report the full models here, including main effects.

tion asker, as mediated by the responsiveness of the question asker. Across 5,000 simulations, this procedure estimated an effect that was significantly different from zero (standardized effect = .12, 95% CI [.01, .25], p = .034). This suggests that responsiveness plays an important role in explaining why asking someone more questions leads to interpersonal liking.

Discussion. In this study, we tested the effect of asking many or few questions in a conversation. When one conversation partner was instructed to ask many (nine) or few (four) questions, unmanipulated conversation partners liked the high question-askers more than they liked the low question-askers. The participants did not anticipate the effect of question-asking on liking. Furthermore, we found evidence for a distinct mechanism underlying the effect. Participants assigned to ask a high number of questions were perceived as more responsive to their partners, which predicted higher liking by their conversation partners.

Study 2

In Studies 2A and 2B, we address two limitations of Study 1. First, in Study 1, we manipulated the question-asking of one of the partners in every dyad. But the other partner was free to choose the number of questions they asked, and could adjust their own question-asking to adapt to their partner's question asking. Thus, we could not make a clean causal test of whether the match or mismatch between the partners' question-asking rates affected their liking of each other. We designed Study 2 to test whether matched or mismatched question-asking would impact liking, or whether the main driver was simply the number of questions asked by one's partner. Second, we only measured liking by the people who were actively involved in the conversation. This meant that we could not determine whether liking was driven by indirect trait inferences (e.g., question-asking serves as an indicator of likability broadly) or by the direct experience of the conversationalists during the conversation itself.

In Study 2, we address these issues directly. Participants in Study 2A again chatted with each another in dyads, but all participants were assigned to either high or low question-asking conditions in a 2 (self: high vs. low question asking) × 2 (partner: high vs. low question asking) design. In Study 2B, we recruited a separate sample of observers to read the transcripts of the conversations in Study 2A and rate both participants on the same dependent measures. These observers were able to take an outside view of the conversation, without having to focus on maintaining the dialogue.

Our analyses throughout Study 2 use hierarchical linear modeling to control for the fact that all of our outcomes are nested within dyads, as we randomized condition at the individual level (as opposed to the dyadic level as in Study 1). This allowed us to estimate the effect of high (vs. low) question-asking instructions in Person A on four hypothesized outcome measures: how much Person B likes A, how much A thinks s/he is liked by B, how much a neutral Observer C likes A, and how much C thinks A is liked by B. Finally, we again tested our full theoretical model by analyzing responsiveness as a mediator of the relationship between question-asking and liking.

Study 2A Method

Participants. We recruited participants from Amazon's Mechanical Turk (MTurk) for a "Chat Study." We recruited a total of 446 participants to target a sample that was the same size as the sample size in Study 1 (N = 430). From that group, we applied the same a priori exclusion criteria as in Study 1 (because the sample was collected online, there were naturally more technical challenges that led to more exclusions compared with Study 1, which was conducted in a behavioral lab). We excluded 15 dyads where at least one participant did not finish the study. We excluded 28 dyads that contained a duplicate IP address and three dyads that contained a duplicate MTurk ID. Finally, we excluded eight dyads in which participants reported that they were not able to complete a full conversation. After exclusions, we analyzed data from 338 participants (177 male, 161 female), or 169 dyads.

Design and procedure. Before starting their conversations, all participants were given the same instructions as in Study 1 and told that their objective was "to get to know each other." The text of the question-asking instructions was also the same (i.e., "ask [at most four/at least nine] questions"). As in Study 1, we measured liking and predicted liking as our main dependent variables (see Appendix A for all measures collected).

The most important difference from Study 1 is that both participants in every conversation were given question-asking instructions (compared with just one participant). Each participant was assigned randomly to ask either many or few questions, and they were assigned to condition independently from their conversation partner. This ensured that in one-quarter of the pairs both participants were assigned to ask many questions, in one quarter of the pairs both participants were assigned to ask few questions, and the remaining pairs included one partner assigned to ask many questions and one partner assigned to ask few. Because the manipulations were at the level of individuals rather than dyads, we could test our effects after controlling for the nested nature of each dyad.

After data collection ended, we recruited four independent raters to code responsiveness for each participant in all 169 conversations in randomized order, using the same measure of responsiveness that we used in Study 1 (Reis et al., 2011).

Study 2A Results

Manipulation check: Questions asked. Across all conditions, individual participants asked 5.98 questions (SD = 3.63), on average. Following the instructions, participants who were told to ask many questions did in fact ask more questions (M = 8.77, SD = 3.15) than did participants who were told to ask few questions (M = 3.52, SD = 1.78; HLM: β = 2.68, SE = .33), $t(184)$ = 8.05, $p < .001$, Cohen's d = 1.45. This difference also held when we computed question-asking as a fraction of total conversational turns: high-question participants had a higher question rate (M = 46.15%, SD = 19.77%) than did low-question participants (M = 22.47%, SD = 15.60%; HLM: β = 12.17%, SE = 1.99%), $t(223)$ = 6.13, $p < .001$, Cohen's d = 1.11. There was no effect of the partner's condition on one's own question rate (all ps > .3). This confirms that our manipulation had its intended effect on how participants conducted their conversations.

Liking. In these data, both participants in each dyad were subject to a manipulation, so we tested our hypotheses using nested hierarchical linear models. Participants reported their liking for their partner using the same four items as in Study 1 (see Appendix A). These items were aggregated into a single standardized index of liking (Cronbach's α = .92), and the results by condition are depicted in Figure 2. The results replicate the effect found in Study 1: participants liked high question-askers (M = 6.02, SD = .74) more than low question-askers (M = 5.79, SD = .97; HLM: β = .28, SE = .09), t(306) = 3.13, p = .001, Cohen's d = .27.

We also conducted a multiple regression model to test for an interaction between experimental conditions. The nonsignificant interaction term revealed that the effect of partner question-asking on liking of partner was not moderated by own question-asking (β = .11, SE = .22), t(166) = .49, p = .628, which implies that the effect of question-asking is robust to variations in question-asking from the person being asked. No matter how many questions you asked your partner, the number of questions s/he asked you influenced your liking of them.

Predicted liking. We tested whether participants anticipated the effects of question-asking, using the same standardized index of predicted liking as in Study 1 (Cronbach's α = .92). Again, participants assigned to ask many questions did not think they would be liked any more (M = 5.16, SD = .90) than participants assigned to ask few questions (M = 5.22, SD = 1.00). Like in Study 1, we again tested a 2 (manipulation: high vs. low question-asking) × 2 (perspective: partner's liking vs. asker's predicted liking) interaction, in a nested model that controlled for the fact that outcomes were nested within dyad. This interaction term was significant (β = −.29, SE = .11, t(504) = 2.55, p = .011 (see the full model in Table 1, Panel B), suggesting that the question-asking manipulation did indeed influence the partners' actual liking more than the askers' prediction of that liking. And there was once again no correlation between question-asking and predicted liking across all conditions (r = −.06), t(336) = 1.18, p = .238. These results provide further evidence that the positive effect of question-asking on liking is not anticipated by the askers.

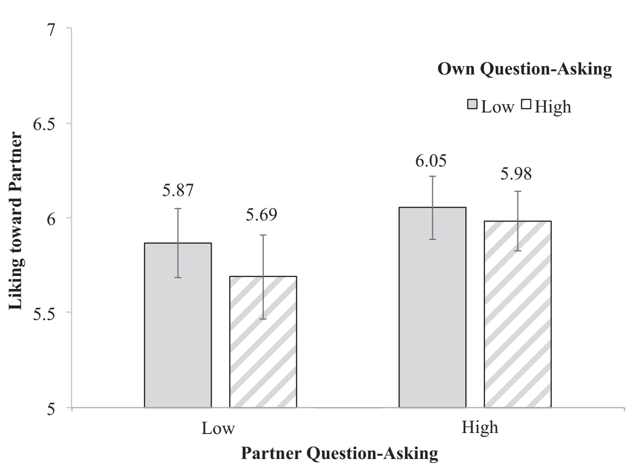

Figure 2. The effect of question-asking on liking in Study 2A. Each participant was given question-asking instructions, such that each person's own instructions and partner's instructions were independently manipulated in a 2 × 2 design. Error bars represent 95% CI for the group means.

Responsiveness. There was once again high agreement among the four coders' ratings of responsiveness (ICC = .75). Replicating the results from Study 1, participants who were instructed to ask many questions were rated as being more responsive to their partner (M = 4.69, SD = .67), compared with participants who were instructed to ask few questions (M = 4.62, SD = .69; HLM: β = .11, SE = .05), t(221) = 2.12, p = .035. We conducted another test of our proposed mediation model, and again found support for our hypotheses. That is, the effect of question-asking instructions on partner liking was significantly mediated by the responsiveness of the question-asker to their partner (standardized effect = .07, 95% CI [.00, .14], p = .041).

Study 2B Method

Participants. We recruited 644 participants from Amazon's Mechanical Turk (MTurk) who participated in exchange for $0.50. As in Studies 1 and 2A, exclusion criteria were determined a priori. We excluded 30 participants with duplicate IP addresses and two participants who reported that they could not read the chat conversation. We included 612 participants (373 male, 239 female) in the analysis.

Design and procedure. Participants were randomly assigned to read the transcript of one of the 169 conversations from Study 2A and were told they would answer some questions about the conversation partners. Afterward, participants reported their own liking of both partners and their prediction of how much each person liked their partner, using the same sets of measures as in Study 2A. These questions were grouped into two blocks—reported liking and predicted liking—and the order of these blocks was counterbalanced. Next, participants reported their estimates of how many questions each partner asked. Importantly, all participants in this study were neither aware of the question-asking manipulations of these conversation partners, nor of the purpose of the original study and our hypotheses.

Study 2B Results

Each of the 169 conversations was viewed by at least three different independent observers. We combined the observers' ratings exactly as in Study 2A: That is, every observer's own liking and predicted liking for both people in the conversation were calculated at the individual level, as a standardized index across the set of four liking items.

In general, observers who rated the same conversations tended to agree with one another, with high intraclass correlations for reported liking (ICC = .81) and predicted liking by each partner toward the other partner (ICC = .88). To test the effect of question-asking on third-party ratings, the observer ratings were entered into a hierarchical linear model, with their rating as the dependent variable, and controlling for rater- and dyad-level nesting. This allowed us to make a precise estimate of how third-party perceptions were influenced by the question-asking instructions.

Third-party liking. Across the conversations, third-party observers reported a mean liking of 5.41 (SD = 1.03) toward participants. When each person's question-asking condition was entered as a predictor in the hierarchical model (controlling for dyad and rater nesting), the results showed that people who were assigned to ask more questions were not liked any more than people who were assigned to ask fewer questions (HLM: $\beta = -.04$, $SE =$ .04), $t(744) = 1.18$, $p = .240$.

The crucial test is whether the question-asking manipulation had a larger effect on the partners' liking than on the third party rater's liking. This required that we test a 2 (manipulation: high vs. low question-asking) × 2 (perspective: partner's own liking vs. third party's liking) interaction, in a nested model that controlled for the fact that outcomes were nested within dyad. The interaction term suggested that the question-asking manipulation influenced the partners' liking more than the observers' liking ($\beta = -.18$, $SE =$ .06), $t(1,669) = 3.25$, $p = .001$ (see the full model in Table 1, Panel C). These results show that question-asking did not induce liking in outside observers in the same way that it affected the person receiving those questions in the conversation.

Though it is not central to our hypotheses, we also tested how *answering* questions affected liking by third-party observers. In our data, outside observers liked people whose partner was instructed to ask many questions more than people whose partner was instructed to ask few questions ($\beta = .08$, $SE = .04$, $t(745) =$ 2.05, $p = .041$). Again, we tested a nested model with a 2 (manipulation: high vs. low questions received) × 2 (perspective: partner's own liking vs. third party's liking) interaction term. The interaction term suggested that the level of questions received influenced the observers' liking more than the partners' liking ($\beta = .22$, $SE = .06$), $t(1,670) = 4.02$, $p < .001$ (see the full model in Table 1, Panel D). These analyses suggest that question-answering, rather than question-asking, makes a person more appealing to third parties.

Third-party predicted liking. Across the conversations, third-party observers reported a mean predicted liking of 5.59 ($SD = .95$) by one partner for the other partner. When each conversation partner's condition was entered as a predictor in the hierarchical model, the results showed that neither one's own question-asking ($\beta = .01$, $SE = .03$), $t(699) = .29$, $p = .773$, nor partner question-asking ($\beta = -.01$, $SE = .03$), $t(699) = .19$, $p =$ .847, influenced third-party predicted liking. These results showed that outside observers, like conversation participants, did not anticipate that people who receive many questions would like their partner more. We once again tested a nested model with a 2 (manipulation: high vs. low question-asking) × 2 (perspective: partner's predicted liking vs. third party's predicted liking) interaction term, and the interaction term suggested that the question-asking manipulation influenced the partner's predicted liking more than the third party's predicted liking ($\beta = -.21$, $SE = .05$), $t(1,670) = 3.95$, $p < .001$ (see the full model in Table 1, Panel E). These results showed that third-party observers, compared to conversation participants, were worse at predicting the effect of question-asking on actual liking.

Discussion

In Study 2A, we again found that question-asking increased liking in a 15-min conversation between two people, replicating the results from Study 1. People who asked a high number of questions were liked by their partner more than people who asked a low number of questions, and this was not moderated by the number of questions the recipient had asked during the conversation. These findings are inconsistent with matching or mismatching hypotheses that might suggest that we like people who ask a similar or different number of questions than we ask. We also replicate our Study 1 results with regard to the underlying psychological mechanism: High question-asking leads to increased responsiveness, which explains the effect of question asking on liking.

Furthermore, in Study 2B, third-party observers did not like high question-askers more than low question-askers. Interestingly, third-party observers liked high question-responders (those who received many questions) more than low question-responders (those who received few questions). In other words, when you are participating in a conversation, you like people who ask more questions. But when you are observing a conversation, you like people who answer more questions. These results suggest that people like question-askers when the questions are directed toward them personally. This further supports the mechanism of responsiveness—we like people who seem responsive to us personally (not to other people in general).

In Studies 2A and 2B, we also replicated our finding that question-askers did not anticipate the effect their questions would have on liking. We extended this result to show that it is also true of third-party observers. That is, even neutral observers did not predict that high question-askers would be liked more by their partners than low question-askers.

Classification of Question Types Across Studies 1 and 2

Studies 1 and 2 provide converging empirical support for our proposed model: people like their partners more when they ask more questions, because people who ask more questions are seen as more responsive. In this section, we analyze not just the number of questions people ask, but also the types of questions they ask across all data from Studies 1 and 2. We investigate several question types descriptively and then focus our predictions and analyses on one question type: follow-up questions (a behavioral indicator of responsiveness). We predict that follow-up questions are an indicator of responsiveness and will correlate positively with liking.

Classification Method

First, we developed a classification scheme for the most common question types in our dataset. On the basis of several small-scale qualitative coding trials, we developed a classification scheme with six question types: follow-up, full-switch, partial-switch, mirror, rhetorical, and introductory (see Table 2 for examples). The first three question types reflected the semantic content of the dialogues, whereas the latter three types captured common structural elements. These question types align with some prior work that categorize questions based on function and type of information sought. For example, follow-up questions reflect general-inquiry questions that begin with why or how, which request that the other person provide information, while rhetorical questions function to make a point rather than request information (Graesser, 1985; Miles, 2013; Schegloff, 2007).

Follow-up questions were those that followed up on the topic the partner had mentioned earlier in the conversation (almost always in the previous turn). One needs to listen and understand what the partner said in order to ask a follow-up question. *Full-*

Table 2
Question Type Examples

Question type	Example
Follow-up	*I'm planning a trip to Canada.* **Oh, cool. Have you ever been there before?**
Full switch	*I am working at a dry cleaners.* **What do you like doing for fun?**
Partial switch	*Not super outdoorsy, but not opposed to a hike or something once in awhile.* **Have you been to the beach much in Boston?**
Mirror	*What did you have for breakfast?* **I had eggs and fruit. How about you?**
Introductory	*Hello!* **Hey, how's it going?**
Rhetorical	*What's the craziest event you've been to?* **Yesterday I followed a marching band around. Where were they going? It's a mystery.**

Note. Examples of conversational turns containing each question type (from Study 1 data). We show the question-asker's turn in boldface type and their partner's previous turn in italic type.

switch questions were those that asked about a new topic, one that was unrelated to what the partner had already discussed. We also included a separate category, *partial-switch questions*, for questions that changed topics somewhat, but not entirely.

Mirror questions describe questions similar in content or structure to a question asked by the partner in a prior turn. Mirror questions are always preceded by a question, and are distinct from follow-up questions, which are preceded by a statement. *Rhetorical questions* were defined as questions where one does not expect a response from the partner; these speech acts take the grammatical form of a question, but are used to make a point rather than elicit information. Finally, *introductory questions* were the most superficial, routine questions at the beginning of the conversation. See Table 2 for examples of each question type drawn from our data.

We trained six research assistants to read through the 368 conversation transcripts from Studies 1 and 2A, such that each transcript would be read by three independent coders. Each person classified every question as one of the six types, and their scores were compiled at the level of every turn. Each turn with a question was assigned a single label based on majority rule among the three coders' labels (ties occurred in <1% of all cases, and were broken in favor of the least prevalent label).

Classification Results and Discussion

Across Studies 1 and 2A, raters had high agreement across most of the question types they were assigned to code. Follow-up questions were the most common, comprising 40.51% of all questions (Cronbach's $\alpha = .87$), followed by full-switch questions (27.55%, $\alpha = .86$), mirror questions (19.03%, $\alpha = .94$), introductory questions (5.52%, $\alpha = .93$), partial-switch questions (5.48%, $\alpha = .47$), and then rhetorical questions (1.91%, $\alpha = .74$). The last two question types were both difficult to define (low alpha) and rare (low prevalence), so we leave out partial-switch and rhetorical questions from the analyses reported below. However, the results for the other question types are unchanged if these question types are included or excluded from the analyses.

We first investigated how the counts of question types mapped onto the constructs in our theoretical model, across all participants in all conditions. We estimated multiple regressions that regressed each of the dependent measures from our model—coder-rated responsiveness and partner-reported liking—onto each participant's set of question counts as independent variables. These results are summarized in Table 3. We find that responsiveness was higher among people who asked more follow-up questions, and lower among people who asked more full-switch questions. We also find that follow-up questions alone predicted increased partner liking.

We then compared question types across our two question-asking conditions and found that not only did people in the high question-asking conditions ask more questions, they asked different kinds of questions. Specifically, the proportion of questions that were follow-up questions was higher among people who were instructed to ask many questions ($M = 41.50\%$, $SD = 20.70\%$) than among people who were instructed to ask few questions ($M = 28.02\%$, $SD = 26.67\%$, HLM: $\beta = .14$, $SE = .02$), $t(526) = 6.61$, $p < .001$. Furthermore, this proportion significantly mediated the effect of condition on partner liking (estimated effect = .06, 95% CI [.01, .11], $p = .006$).

Though we instructed participants to ask more versus fewer questions, we did not provide guidance about which question types to ask. So why did so many of the "extra" questions become follow-up questions? We speculate that this phenomenon arose because high question-askers are likely to draw out more information from (and learn more about) their partners. This information is salient and immediately accessible during the conversation, and makes it easy to ask follow-up questions that continue on the topic at hand. Furthermore, follow-up questions from participants in all conditions were more common in turns later in the conversation, as more information was shared and rapport was developed. Follow-up questions are an easy and effective way to keep the conversation going, and show that the asker has paid attention to what their partner has said.

Our analyses of question types suggest a specific prescription for conversationalists. Asking more questions broadly leads to more interpersonal liking, but follow-up questions are particularly likely to increase liking because they require responsiveness from the question-asker, and signal responsiveness to the question-asker's partner.

Table 3
Multiple Regressions of Question Type Rates on Responsiveness and Partner Liking (Studies 1 & 2)

Question type	Responsiveness		Partner liking	
	M	*SE*	*M*	*SE*
Follow-up question rate	1.92^{***}	.35	$.89^{*}$	.36
Full switch question rate	-2.83^{***}	.41	−.29	.43
Mirror question rate	.16	.47	−.03	.49
Introductory question rate	.98	1.11	−.80	1.17
Sample size	368		368	

Note. Multiple regressions are on two dependent measures: the asker's responsiveness to their partner and how much that partner likes the asker. $^{\wedge}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .005$.

Study 3

In Studies 1 and 2, we found that question-asking increases responsiveness and liking, and these effects were driven by increases in follow-up questions, rather than other types of questions. In Study 3, we build on these results by examining the natural relationship between question-asking and a behavioral measure of liking—agreeing to a second date—in speed-dating events.

In this study, we use data from a previous investigation of speed-dating (Jurafsky, Ranganath, & McFarland, 2009; Ranganath, Jurafsky, & McFarland, 2009, 2013). Speed-dating is an ideal environment to test our hypotheses because (a) speed daters are highly motivated to make a positive first impression on their conversational partners and (b) each speed dater interacts with many other partners, which allows us to estimate individual differences in question-asking. Furthermore, the behavioral outcome of second-date agreement allows us to estimate the real relationship between question asking and liking, and quantify that relationship relative to the influences of other variables observable in this setting (e.g., gender).

Method

One hundred ten men and women were gathered in three different speed-dating sessions. Each dater went on 15 to 19, 4-min speed dates during a session. Every person wore a microphone to capture the dialogue during the dates. After each speed date, participants filled out a brief survey about their most recent partner, and indicated whether they would want a follow-up date. If both people in a pair wanted a second date, the experimenters provided them with each other's contact information (see McFarland, Jurafsky, & Rawlings, 2013 for a full description of methods).

Our primary dependent variable was participants' willingness to go on a second date, recorded as a binary response (yes/no). Our primary independent variable was the number of questions each person asked on each date. As in previous studies, we measured question-asking by counting conversational turns that included a question mark. However, in this study, we excluded "repair questions" (asking someone to repeat themselves because they were not heard correctly). These did not exist in the text-based conversations in Studies 1 through 3, but they did constitute 1.31% of all questions in Study 3, according to the detection algorithm reported by Ranganath and colleagues (2013).

On the basis of the results of our question type coding in Studies 1 and 2, we wanted to again test the role of follow-up questions in this dataset. However, the size of the text corpus (n = 987) was much larger than in Studies 1 and 2 (n = 368 combined). So instead of using human coders, we leveraged the coded data from the earlier studies to build an algorithm that could classify question types directly in the Study 3 text data.

The coded data from the first two studies contained 4,209 conversational turns that were classified by human coders to contain one of the six question types described earlier. We processed the text of those questions to create a dataset that contained counts of common words and phrases, as well as some holistic features like turn length and questions in the previous turn (full details in Appendix B). This was used as training data for a LASSO classification algorithm (Friedman, Hastie, & Tibshirani, 2010; Tibshirani, 1996), which could classify new questions based on how their text matched examples from the training data. When we conducted out-of-sample tests using hand-labeled data, the algorithm achieved an 86.97% accuracy rate detecting follow-up questions.

In Study 3, we used the algorithm to label the questions extracted from each speed dater's transcript. The new question turns were processed in the same way as the turns in the training data, to extract the same set of features. Then, each question was assigned four probabilities, based on the predicted likelihood that it was either a follow-up, full-switch, introductory, or mirror question. Our analyses use these probabilities directly to account for model uncertainty. However, our analyses are similar if we instead assign a single label to each new question turn.

Results

Our analyses focus on the 1,961 unique observations for which we had a complete transcript matched to a post-date survey (filled out after the speed-date). All of the dates were filled with active conversation. Though these speed dates were shorter in duration than the conversations in Studies 1 and 2 (4 min compared with 15 min), they produced a similar amount of dialogue, perhaps because talking is a more efficient medium than instant-messaging.

Although the dataset for Study 3 is large and rich, it is observational, and thus we face two empirical challenges in estimating the relationship between question-asking and liking. The first challenge is that the variation in people's question-asking behavior is endogenous. That is, each speed-dater is free to choose their question-asking rate on the basis of personal and situational factors, including aspects of their partner in that particular date. That means that any relationships we estimate will be correlational. Furthermore, that correlation could be at the trait level (i.e., stable question-asking behavior in individuals across dates) or at the date level (i.e., changes from an individual's baseline question-asking behavior within each date). Our data allow us to test for both types of correlations in fixed effects regressions, using a much larger sample than the "no instructions" participants in Study 1.

The second challenge we face with this dataset is that daters vary in their pickiness. That is, they vary in their general willingness to go on a second date, across all partners. Fortunately, daters had no control over their sequence of partners, so we can easily control for partner-level pickiness. We control for this pickiness in two sets of model specifications. The first set includes gender as a covariate (which accounts for most of the variance in partner-level pickiness across the entire sample), and the second set estimates partner-level pickiness directly.

Question-asking. Questions were common. On an average date, each person asked 9.80 questions (SD = 5.30), which means that roughly 22% of their conversational turns included a question. Our algorithm estimated that the average date included 4.51 follow-up questions from each person (SD = 2.90), 3.20 full-switch questions (SD = 1.80), .33 introductory questions (SD = .37), and 1.76 mirror questions (SD = 1.09).

Question-asking was a relatively stable trait across individuals. That is, the best predictor of someone's question rate on any given speed date was their average question rate on their other dates (r = .53), $t(1{,}959) = 28.02$, $p < .001$. Likewise, the best predictor of someone's follow-up question rate was their average follow-up question rate on other dates (r = .56), $t(1{,}959) = 29.61$, $p < .001$.

These results suggest that most of the meaningful variation in question-asking behavior is at the level of individual daters, not at the level of specific dates.

Liking. Across these observations, 46.81% of participants were willing to go on a second date with their partner. There was a large gender gap in second-date agreement: men wanted second dates with 56.78% of their partners, while women only wanted second dates with 36.84% of their partners, on average. We use these date "success rates" as an indicator of liking. In Figure 3, we plot the average follow-up question rate of every speed dater ($N =$ 110) against the success rate of every speed dater, controlling for baseline gender disparities in success rate (this adjustment is identical to the regression model in column 6 of Table 4). The plot shows a positive relationship: people who ask more follow-up questions find that more of their partners want to go on second dates with them.

We estimated a series of logistic regression models to formally test the relationship between the extent of a person's question-asking and the chances that their partner will want a second date with them (see Table 4). We treat every yes/no second-date decision as an observed outcome ($N = 1961$), however we adjust all our standard errors to account for the fact that these observations are not independent. Specifically, both the outcomes ("yes/no" second date decisions) and the predictors (number of questions asked) are correlated within person, which reduces the effective degrees of freedom in these observations and biases our standard errors toward zero. For this reason, we correct our standard errors to account for this two-way clustering (Cameron, Gelbach, & Miller, 2011), such that our 1961 observations are clustered at the level of both question-asker and question-recipient.

We operationalize our independent variable (question-asking) in three different ways: total questions per date (in columns 1 and 4), total questions per turn (in columns 2 and 5), and question types per turn (in columns 3 and 6). We also report models with different control variables. In some models, we control for gender differences in willingness to date, which account for a large part of the residual variance in date success (columns 4 through 6).

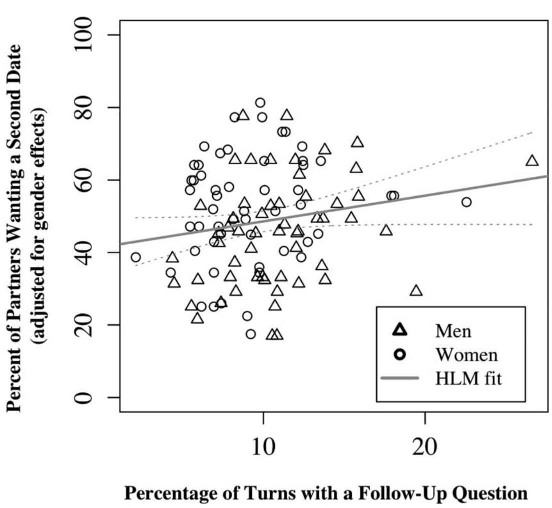

Figure 3. The effect of question-asking on dating outcomes in Study 3. Each point is one speed dater. The *x* axis represents the average number of follow-up questions they asked per date (as a percentage of conversational turns), whereas the *y* axis shows each person's speed dating success rate (i.e., getting second-date requests from their partners). The success rate is adjusted for the gender gap in success rates (women are much pickier, on average). The gray line represents the regression fit on the adjusted data (error bands indicate 95% CI of regression slope).

The results show that the follow-up question rate predicts date success: people who ask a higher proportion of follow-up questions have increased date success, even though other types of question-asking are unrelated to date success. This result was orthogonal to gender, which was the best predictor of success across all daters. Furthermore, total question-asking marginally predicts date success. In additional tests, we found no interaction between gender and question-asking. That is, question-asking was similarly related to date success for both men and women.

We also report another set of regression models that include person-level fixed effects, to control for person-level sources of variation in the speed dater's behavior. Specifically, we use rater-level fixed effects to control for person-level variation in the pickiness of partners (Table 5, columns 1 through 6). Additionally, some models also use asker-level fixed effects to control for person-level variation in the desirability of askers (Table 5, columns 4 through 6). We find that accounting for partner-level pickiness, consistent with our gender controls, simply captures residual variance in date success that is unrelated to the effect of question-asking. However, controlling for asker-level desirability reduces the independent influence of question-asking on date success. Along with the above results on individual-level stability of question-asking behavior, this suggests that our data support a trait-level model of question-asking behavior in this context.

Discussion

The results of Study 3 demonstrate a similar relationship as in Studies 1 and 2 in a face-to-face setting: speed dating. People who asked more follow-up questions were asked on more second dates, a straightforward indicator of interpersonal liking. The regression estimates imply that an 8% difference in follow-up question rates in this sample is associated, on average, with success on one additional date over an evening of 20 speed dates. Of course, one cannot interpret this relationship causally. But Study 3 provides correlational evidence from a field setting that is consistent with the experimental evidence from Studies 1 and 2 that one's follow-up question rate predicts liking from one's partner.

General Discussion

Conversations are complex social interactions, fraught with decisions about what to say and how to behave. Although most adults have decades of experience conversing with others, our data suggest that people often fail to engage in behaviors that will help them make the most positive impression. Whereas prior data demonstrate that people tend to talk about themselves (Dunbar et al., 1997; Marr & Cable, 2014), our results suggest this may not be an optimal strategy. Instead, across several studies, we find a positive relationship between question-asking and liking. Furthermore, we identify an important psychological mechanism: the effect of question-asking is driven by an increase in perceived responsiveness, which leads question-recipients to like their partners more. Across our studies, we find support for this mechanism

Table 4
Regression Models from Study 3

Predictor variable	1	2	3	4	5	6
Total question count	.10 (.09)			.15^ (.09)		
Total question rate		.08 (.08)			.14^ (.08)	
Full-switch question rate			−.12 (.11)			−.08 (.11)
Follow-up question rate			.18^ (.10)			.24* (.10)
Introductory question rate			.04 (.05)			.09 (.05)
Mirror question rate			.00 (.07)			−.06 (.07)
Gender (1 = Male)				.85*** (.24)	.85*** (.24)	.89*** (.24)
pseudo-R^2	.002	.001	.002	.033	.032	.035

Note. Observations: $N = 1{,}961$. Each column represents a different logistic regression model from the data in Study 3. Each model estimates the relationship between a speed-dater's question-asking behavior and their partners' willingness to go on a second date. Values in parentheses indicate standard errors. All standard errors are adjusted to be two-way robust to correlations within askers and within their partners. Models 1 through 3 test the simple relationships between question-asking and partner liking, and Models 4 through 6 control for the large disparity in date success by gender.
^ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .005$.

using both an attitudinal and a behavioral measure of responsiveness. In particular, we identify follow-up questions as an important behavioral indicator of responsiveness, and we find that asking a higher rate of follow-up questions reliably predicts partner liking.

By asking third-party observers to rate conversations, we tested whether question askers are liked because they are more responsive, or because question-asking is a signal of traits that are inherently likable—for example by signaling interest, empathy, or superior social skills. Instead, we found that high question-askers were not liked any more by third-party observers than were low question-askers. This result dovetails with our findings on follow-up questions, perhaps suggesting a boundary condition. Because a third-party observer is not present in the conversation by definition, none of the questions being asked can follow up on anything they have said. These results provide converging evidence that people like question-askers because they perceive question-askers as more responsive (to them personally).

In this research, we find evidence that people do not anticipate the effect of question-asking on liking. This is true when they are in the role of question-asker, question-receiver, and third-party observer. Thus, it seems that people are largely unaware that asking questions has social benefits. Indeed, third-party raters seemed to like question-responders more, perhaps because they were interested in the information that responses provide. This result may explain why low question-asking persists in social interaction: individuals may recall that while observing conversation, they liked or found interesting high question-responders.

Last, we investigated conversations in the ecologically valid context of speed-dating, where people are motivated to make a good impression during a first encounter. We found that people who asked a higher rate of follow-up questions were asked on more second dates. Using a machine learning analysis of the question contents, we found a positive relationship between follow-up questions and liking in this context, but not between other types of questions and liking. The findings from Study 3 provide correlational evidence supporting the experimental evidence from Studies 1 and 2 that follow-up questions, as a behavioral measure of responsiveness, are particularly likely to increase liking. Furthermore, we measured question-asking across many interactions for each person, and found evidence that asking

Table 5
Regression Models Including Partner-Level Effects from Study 3

Predictor variable	1	2	3	4	5	6
Total question count	.30*** (.11)			.07 (.12)		
Total question rate		.17^ (.10)			−.20 (.13)	
Full-switch question rate			−.13 (.12)			−.28 (.19)
Follow-up question rate			.31* (.13)			.11 (.20)
Introductory question rate			.07 (.07)			−.10 (.08)
Mirror question rate			−.06 (.09)			−.02 (.09)
Partner fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Asker fixed effects	No	No	No	Yes	Yes	Yes
pseudo-R^2	.268	.262	.265	.432	.433	.435

Note. Observations: $N = 1{,}961$. Each column represents a different logistic regression model from the data in Study 3. Each model estimates the relationship between a speed-dater's question-asking behavior and their partners' willingness to go on a second date. All standard errors (in parentheses) are adjusted to be two-way robust to correlations within askers and within their partners. Additionally, Models 1 through 3 control for partner fixed effects (i.e. overall pickiness of partners), and Models 4 through 6 account for asker fixed effects (i.e. overall desirability of askers).
^ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .005$.

follow-up questions is a relatively stable trait over time. The results from Study 3 suggest that follow-up question-asking is a desirable trait that people may seek in potential partners.

Theoretical Implications

Our work makes several fundamental theoretical contributions to an array of existing literatures. First, our work contributes to the understanding of responsiveness within the context of conversations. When Person A asks Person B more questions, particularly follow-up questions, Person B will like Person A more as a result. People want to be heard and validated by others (Davis & Perkowitz, 1979; Laurenceau et al., 1998; Reis & Patrick, 1996; Reis & Shaver, 1988). The question-asker, by expressing interest and engagement, serves a validating role as a valuable conversation partner—indeed, one with whom people want to interact in the future. Prior research has conceptualized responsiveness as understanding, validation, and care (Reis & Patrick, 1996; Reis & Shaver, 1988), and we show that an important behavioral indicator of responsiveness is asking more follow-up questions in a conversation. Follow-up questions appropriately elaborate on the content of the partner's message, and signal that the content is worth continuing to discuss. Responsiveness has previously been conceptualized in close relationships, and we build on recent work that studies responsiveness in casual encounters and first meetings (e.g., Reis et al., 2011), by identifying a behavior that can signal responsiveness during conversations.

Second, our work contributes to extant research about active listening, which has been previously investigated across fields such as communications, crisis communication, and marital therapy (e.g., Bodie, 2011; Bodie et al., 2012; Gordon, 1975; Lester, 2002; Rogers, 1951; Stanley, Bradbury, & Markman, 2000; Weger et al., 2014) but has been largely overlooked in social psychology. We identify and show evidence that question-asking is a critical component of active listening. Bridging the literatures of communications and social psychology, we suggest that question-asking is an important indicator of both active listening and responsiveness, and we open the pathway for future research to investigate active listening as a social psychological construct.

Third, our finding that people fail to predict the effect of question-asking on liking contributes to previous work on prediction and forecasting errors (e.g., Gilbert & Wilson, 2007). We suspect that people may show a truncation error specific to conversational experiences: When people simulate conversations, they tend to only imagine what they would say at one time point, rather than the timing and responsiveness of dialogue as the conversation unfolds. This prediction error could additionally help to explain why people tend to talk about themselves during a conversation rather than ask many questions. At any one time point, it is easier to offer statements about the self, since that information is more easily accessible compared to responses that are contingent on the partner's response. Thus, people fail to predict the effect of question-asking on liking when reflecting on conversation and when engaged in a conversation.

Fourth, these findings contribute to a literature about interpersonal interaction and intimacy in longer-term relationships. The process of interacting with another person affects outcomes such as attraction and intimacy in relationships (Davis & Perkowitz, 1979; McAllister, 1980; Miller et al., 1983; Reis, Clark, & Holmes, 2004; Reis & Shaver, 1988). Although our work only looked at first-encounters, it is likely that the effect of question-asking on responsiveness and liking extends to repeated interactions and longer term relationships. Previous work has found that social closeness increases as people reciprocally answer questions that grow increasingly more intimate (Aron et al., 1997; Sedikides et al., 1999). But this work has focused on full-switch questions, provided in advance by an experimenter. Our results focus on follow-up questions in natural conversation, and suggest that follow-up questions, as an indicator of responsiveness, may be an important factor for attraction and intimacy in longer term relationships.

Finally, even though question-asking seems to be a relatively stable trait across individuals, people can learn to ask more questions. In Studies 1 and 2, we manipulated question-asking, showing that it is remarkably easy to induce people to ask more questions. This suggests that question-asking is a skill that can be learned. For example, if a person consistently receives positive feedback (such as liking from the other person, or more second dates) for asking more questions during a conversation, she could learn to associate question-asking with positive outcomes, and change her conversational behavior. Over time, given a choice to talk about oneself or ask more questions in a conversation, a person could tend to choose the latter. This is consistent with work showing that active listening skills can be improved with training (Paukert et al., 2004).

Limitations and Future Directions

Our findings are qualified by several limitations that suggest fruitful avenues for future research. For example, in our experiments, we randomly induced the decision to ask questions in conversation. However, in the real world, this decision is not exogenous and may be influenced by many contextual or person-level factors. In our speed-dating study, we find evidence that question-asking is a relatively stable trait across individuals. This result is consistent with prior work showing that being an "opener" in conversation is a stable trait (Miller et al., 1983). Although context was controlled in our experiments, we also found evidence for one important contextual factor: time. Specifically, in all of our studies, people asked more questions early in their conversation, and the question rate drifted downward throughout (though the rate of speech remained constant). However, the number of follow-up questions increased over time, even as other question types became much less common. Of course, time is not the only contextual driver of question-asking, and more research is needed.

Next, the relationship between question-asking and interpersonal liking may not be monotonic—it may be curvilinear. That is, it may be the case that one can ask too many questions, annoying the conversation partner while not revealing sufficient information about oneself. When one asks too many questions without reciprocating self-disclosure, liking may decrease. People do not want to talk about themselves indefinitely without reciprocity, since self-disclosure should induce the other person to self-disclose as well (Cozby, 1973; Dindia, 1988, 2002; Jourard, 1971; Sprecher et al., 2013). Indeed, prior research suggests that higher turn-taking in reciprocal self-disclosure increases liking (Sprecher et al., 2013). A person who asks too many questions would likely be contributing less to the content of the conversation, by not presenting enough challenge or novelty (e.g., Silvia, 2008). In our data, third-party observers preferred question-recipients, perhaps because those who answered many questions seemed like more complete, interesting people. Therefore, there is likely an optimal balance between question-asking and question-answering. To test this idea, future work should test question-asking dynamics at the extremes. What happens when someone asks zero questions? What happens when someone asks 50?

Broadly, our work opens a new research topic and methodology on the psychology of conversation, a pervasive human experience. We contribute a new methodology for studying conversations using software (ChatPlat) that facilitates and captures the text of live conversations. We conducted a set of large-sample experimental studies involving live dyadic interaction, which is unprecedented. The use of these methods is important for studying live dyadic interactions between people, since previous research has been limited to small sample sizes of dyads, hypothetical scenarios, and confederates. Furthermore, we show how basic features of open-ended conversations can be classified automatically, allowing precise measurement of experimental data without imposing artificial constraints on the participants themselves. We used natural language processing and machine learning algorithms to build a "follow-up question detector" (see Appendix B) that can be applied to any text data to more deeply understand question-asking dynamics.

Our analyses demonstrate a clear effect of asking follow-up questions, which suggests the importance of question type in conversation. Our results on follow-up questions contribute to a more comprehensive question typology (Graesser, 1985; Miles, 2013; Schegloff, 2007), which may include disclosure triggers that function like questions (e.g., "Tell me more"). Although we have identified that follow-up questions increase liking, there are certainly some types of questions that would not engender liking. For example, rude questions could degrade the quality of a conversation. But rude statements can be equally jarring, so it is not clear whether this is a moderator of question-asking per se or a simple effect of semantic content (i.e., rude vs. polite). Similarly, one might expect questions that would embarrass the question-responder to decrease liking. The effects of question-asking on liking are likely to be moderated by the social sensitivity of the questions being asked. We leave ideas about question type as a fruitful avenue for future research.

Furthermore, question-asking is likely to function differently in competitive contexts. People are more likely to be skeptical or defensive when asked questions that solicit information. For example, people may reply honestly or dishonestly depending on the type of question asked (Minson, Ruedy, & Schweitzer, 2011). Indeed, they may require the question-asker to mutually disclose information in order to feel comfortable disclosing. The reciprocity of disclosure becomes more salient and important in competitive compared to cooperative contexts.

Although there are benefits of question-asking, our findings suggest that people fail to ask enough questions. Why might people forego asking questions in dyadic conversation? First, given that people tend to be egocentric, they may not think to ask questions of their conversation partners at all because they are too focused on their own thoughts, feelings, and beliefs (e.g., Gilovich et al., 2000). Second, people may realize they can ask questions but have too little interest, care, or curiosity to hear the answers. They may lack social curiosity, defined as "interest in how other people behave, think, and feel" (Renner, 2006, p. 305). Third, people may want to ask questions but perceive social costs to asking questions. For instance, asking for help feels awkward and embarrassing (Bohns & Flynn, 2010; DePaulo & Fisher, 1980). People fear appearing incompetent when asking others for advice (Brooks, Gino, & Schweitzer, 2015) and may hesitate to ask questions because they fear rejection from others (Downey & Feldman, 1996). In the classroom, students could worry that question-asking shows ignorance (Graesser, McMahen, & Johnson, 1994). Indeed, when placed in private tutoring settings compared to the public arena of the classroom, student question-asking increases to 10 questions per hour (Graesser, Person, & Huber, 1993). Furthermore, people may worry about making a negative impression by asking the "wrong" questions—those that could be perceived as rude, inappropriate, or intrusive. In sum, people may underweight the benefits of question-asking, overweight the costs, or both.

Conclusion

People spend most of their time during conversations talking about their own viewpoints and tend to self-promote when meeting people for the first time. In contrast, high question-askers—those that probe for information from others—are perceived as more responsive and are better liked. Although most people do not anticipate the benefits of question-asking and do not ask enough questions, people would do well to learn that it doesn't hurt to ask.

References

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63,* 596–612. http://dx.doi.org/10.1037/0022-3514.63.4.596

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology, 60,* 241–253. http://dx.doi.org/10.1037/0022-3514.60.2.241

Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin, 23,* 363–377. http://dx.doi.org/10.1177/0146167297234003

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(7).

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin, 91,* 3–26. http://dx.doi.org/10.1037/0033-2909.91.1.3

Berger, C. R., & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research, 1,* 99–112. http://dx.doi.org/10.1111/j.1468-2958.1975.tb00258.x

Berman, J. Z., Levine, E. E., Barasch, A., & Small, D. A. (2015). The braggart's dilemma: On the social rewards and penalties of advertising prosocial behavior. *Journal of Marketing Research, 52,* 90–104. http://dx.doi.org/10.1509/jmr.14.0002

Berscheid, E. (1985). Interpersonal attraction. In G. Lindzey & K. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 413–484). New York, NY: Random House.

Berscheid, E., & Regan, P. (2005). *The psychology of interpersonal relationships*. Upper Saddle River, NJ: Prentice Hall.

Bertakis, K. D., Roter, D., & Putnam, S. M. (1991). The relationship of physician medical interview style to patient satisfaction. *The Journal of Family Practice, 32,* 175–181.

Bodie, G. D. (2011). The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly, 59,* 277–295. http://dx.doi.org/10.1080/01463373.2011.583495

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening, 26,* 1–28. http://dx.doi.org/10.1080/10904018.2012.639645

Bohns, V., & Flynn, F. J. (2010). "Why didn't you just ask?" Underestimating the discomfort of help-seeking. *Journal of Experimental Social Psychology, 46,* 402–409. http://dx.doi.org/10.1016/j.jesp.2009.12.015

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science, 61,* 1421–1435. http://dx.doi.org/10.1287/mnsc.2014.2054

Brooks, A. W., & Schweitzer, M. E. (2011). Can Nervous Nelly negotiate? How anxiety causes negotiators to make low first offers, exit early, and earn less profit. *Organizational Behavior and Human Decision Processes, 115,* 43–54. http://dx.doi.org/10.1016/j.obhdp.2011.01.008

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics, 29,* 238–249. http://dx.doi.org/10.1198/jbes.2010.07136

Chafe, W. L. (1970). *Meaning and the structure of language*. Chicago, IL: University of Chicago Press.

Chen, F. S., Minson, J. A., & Tormala, Z. L. (2010). Tell me more: The effects of expressed interest on receptiveness during dialog. *Journal of Experimental Social Psychology, 46,* 850–853. http://dx.doi.org/10.1016/j.jesp.2010.04.012

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13,* 259–294. http://dx.doi.org/10.1207/s15516709cog1302_7

Cozby, P. C. (1973). Self-disclosure: A literature review. *Psychological Bulletin, 79,* 73–91. http://dx.doi.org/10.1037/h0033950

Davis, D. (1982). Determinants of responsiveness in dyadic interaction. In W. Ickes & E. S. Knowles (Eds.), *Personality, roles, and social behavior* (pp. 85–139). New York, NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4613-9469-3_4

Davis, D., & Perkowitz, W. T. (1979). Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction. *Journal of Personality and Social Psychology, 37,* 534–550. http://dx.doi.org/10.1037/0022-3514.37.4.534

DePaulo, B., & Fisher, J. (1980). The costs of asking for help. *Basic and Applied Social Psychology, 1,* 23–35. http://dx.doi.org/10.1207/s15324834basp0101_3

Dillon, J. T. (1982). The multidisciplinary study of questioning. *Journal of Educational Psychology, 74,* 147–161. http://dx.doi.org/10.1037/0022-0663.74.2.147

Dillon, J. T. (1988). *Questioning and teaching: A manual of practice*. New York, NY: Teachers College Press.

Dindia, K. (1988). A comparison of several statistical tests of reciprocity of self-disclosure. *Communication Research, 15,* 726–752. http://dx.doi.org/10.1177/009365088015006004

Dindia, K. (2002). Self-disclosure research: Knowledge through meta-analyses. In M. Allen, R. W. Preiss, B. M. Gagle, & N. Burrell (Eds.), *Interpersonal communication: Advances through meta-analyses* (pp. 169–186). Mahwah, NJ: Lawrence Erlbaum.

Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology, 70,* 1327–1343. http://dx.doi.org/10.1037/0022-3514.70.6.1327

Dunbar, R. I., Marriott, A., & Duncan, N. D. (1997). Human conversational behavior. *Human Nature, 8,* 231–246. http://dx.doi.org/10.1007/BF02912493

Duncan, S. D., Jr., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory*. Hillsdale, NJ: Lawrence Erlbaum.

Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 498–541). New York, NY: Wiley http://dx.doi.org/10.1002/9780470561119.socpsy001014

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software, 25,* 1–54. http://dx.doi.org/10.18637/jss.v025.i05

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33,* 1–22. http://dx.doi.org/10.18637/jss.v033.i01

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science, 317,* 1351–1354. http://dx.doi.org/10.1126/science.1144161

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology, 78,* 211–222. http://dx.doi.org/10.1037/0022-3514.78.2.211

Godfrey, D. K., Jones, E. E., & Lord, C. G. (1986). Self-promotion is not ingratiating. *Journal of Personality and Social Psychology, 50,* 106–115. http://dx.doi.org/10.1037/0022-3514.50.1.106

Gordon, T. (1975). *P. E. T.: Parent effectiveness training*. New York, NY: New American Library.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25,* 161–178. http://dx.doi.org/10.2307/2092623

Graesser, A. C. (1985). An introduction to the study of questioning. In A. C. Graesser & J. B. Black (Eds.), *The psychology of questions* (pp. 1–14). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., McMahen, C. L., & Johnson, B. (1994). Question asking and answering. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.

Graesser, A. C., Person, N. K., & Huber, J. (1993). Question asking during tutoring and the design of educational software. In M. Rabinowitz (Ed.), *Cognitive science foundations for instruction* (pp. 149–172). Hillsdale, NJ: Lawrence Erlbaum.

Graham, N., Arai, M., & Hagströmer, B. (2016). *multiwayvcov: Multi-way standard error clustering*. R package version 1.2.3. Retrieved from https://cran.r-project.org/web/packages/multiwayvcov/multiwayvcov.pdf

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York, NY: Academic Press.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis, 21,* 298–313. http://dx.doi.org/10.1093/pan/mps028

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107,* 65–81. http://dx.doi.org/10.1037/0033-2909.107.1.65

Hutchby, I. (2005). "Active listening": Formulations and the elicitation of feelings-talk in child counseling. *Research on Language and Social Interaction, 38,* 303–329. http://dx.doi.org/10.1207/s15327973rlsi3803_4

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology, 99,* 549–571. http://dx.doi.org/10.1037/a0020386

John, L. K., Barasz, K., & Norton, M. I. (2016). Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences the United States of America, 113,* 954–959. http://dx.doi.org/10.1073/pnas.1516868113

Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 1, pp. 231–262). Hillsdale, NJ: Lawrence Erlbaum.

Jourard, S. M. (1971). *The transparent self* (rev. ed.). New York, NY: Van Nostrand Reinhold.

Jurafsky, D., & Martin, J. (2009). *Speech and natural language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.

Jurafsky, D., Ranganath, R., & McFarland, D. (2009, May). Extracting social meaning: Identifying interactional style in spoken conversation. In

Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American chap. of the Association for Computational Linguistics (pp. 638–646). Association for Computational Linguistics, Boulder, Colorado.

Kearsley, G. P. (1976). Questions and question asking in verbal discourse: A cross-disciplinary review. *Journal of Psycholinguistic Research, 5,* 355–375. http://dx.doi.org/10.1007/BF01079934

Landis, M. H., & Burtt, H. E. (1924). A Study of Conversations. *Journal of Comparative Psychology, 4,* 81–89. http://dx.doi.org/10.1037/h0071502

Laurenceau, J. P., Barrett, L. F., & Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: The importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of Personality and Social Psychology, 74,* 1238–1251. http://dx.doi.org/10.1037/0022-3514.74.5.1238

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107,* 34–47. http://dx.doi.org/10.1037/0033-2909.107.1.34

Lester, D. (2002). Active listening. In D. Lester (Ed.), *Crisis intervention and counseling by telephone* (2nd ed., pp. 92–98). Springfield, IL: Charles C Thomas.

Marr, J. C., & Cable, D. M. (2014). Do interviewers sell themselves short? The effects of selling orientation on interviewers' judgments. *Academy of Management Journal, 57,* 624–651. http://dx.doi.org/10.5465/amj.2011.0504

McFarland, D. A., Jurafsky, D., & Rawlings, C. M. (2013). Making the connection: Social bonding in courtship situations. *American Journal of Sociology, 118,* 1596–1649.

McAllister, H. A. (1980). Self-disclosure and liking: Effects for senders and receivers. *Journal of Personality, 48,* 409–418. http://dx.doi.org/10.1111/j.1467-6494.1980.tb02376.x

Miles, E. W. (2013). Developing strategies for asking questions in negotiation. *Negotiation Journal, 29,* 383–412. http://dx.doi.org/10.1111/nejo.12034

Miller, L. C., & Berg, J. (1984). Selectivity and urgency in interpersonal exchange. In V. J. Derlega (Ed.), *Communication, intimacy, and close relationships* (pp. 161–206). Orlando, FL: Academic Press. http://dx.doi.org/10.1016/B978-0-12-210840-2.50012-6

Miller, L. C., Berg, J. H., & Archer, R. L. (1983). Openers: Individuals who elicit intimate self-disclosure. *Journal of Personality and Social Psychology, 44,* 1234–1244. http://dx.doi.org/10.1037/0022-3514.44.6.1234

Minkin, N., Braukmann, C. J., Minkin, B. L., Timbers, G. D., Timbers, B. J., Fixsen, D. L., . . . Wolf, M. M. (1976). The social validation and training of conversational skills. *Journal of Applied Behavior Analysis, 9,* 127–139. http://dx.doi.org/10.1901/jaba.1976.9-127

Minson, J. A., Ruedy, N. E., & Schweitzer, M. E. (2011, August). *Ask (the right way) and you shall receive: The effect of question type on information disclosure and deception*. Paper presented at the Academy of Management Meeting, San Antonio, TX.

Montoya, R. M., & Insko, C. A. (2008). Toward a more complete understanding of the reciprocity of liking effect. *European Journal of Social Psychology, 38,* 477–498. http://dx.doi.org/10.1002/ejsp.431

Notarius, C., & Markman, H. (l981). Couples interaction scoring system. In E. Filsinger & R. Lewis (Eds.), *Observing marriage: New behavioral approaches*. Beverly Hills, CA: SAGE.

Patrick, B. C., & Reis, H. T. (1995). *Responsiveness to self-disclosure as a mechanism for intimacy development: The relative effects of interpersonal understanding and validation*. (Unpublished manuscript). Rochester, NY: Department of Clinical & Social Sciences in Psychology, University of Rochester.

Patterson, M. L. (1994). Strategic functions of nonverbal exchange. In J. A. Daly & J. M. Wiemann (Eds.), *Strategic interpersonal communication* (pp. 273–293). Hillsdale, NJ: Lawrence Erlbaum.

Paukert, A., Stagner, B., & Hope, K. (2004). The assessment of active listening skills in helpline volunteers. *Stress, Trauma and Crisis, 7,* 61–76. http://dx.doi.org/10.1080/15434610490281075

Ranganath, R., Jurafsky, D., & McFarland, D. (2009). It's not you, it's me: Detecting flirting and its misperception in speed-dates. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Vol. 1* (pp. 334–342). Association for Computational Linguistics, Singapore.

Ranganath, R., Jurafsky, D., & McFarland, D. A. (2013). Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates. *Computer Speech & Language, 27,* 89–115. http://dx.doi.org/10.1016/j.csl.2012.01.005

Reis, H. T., Clark, M. S., & Holmes, J. G. (2004). Perceived partner responsiveness as an organizing construct in the study of intimacy and closeness. In D. Mashek & A. P. Aron (Eds.), *Handbook of closeness and intimacy* (pp. 201–225). Mahwah, NJ: Lawrence Erlbaum.

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology, 101,* 557–570. http://dx.doi.org/10.1037/a0022885

Reis, H. T., & Patrick, B. C. (1996). Attachment and intimacy: Component processes. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 523–563). New York, NY: Guilford Press.

Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process. In S. Duck & D. F. Hay (Eds.), *Handbook of personal relationships: Theory, research, and interventions* (pp. 367–389). Chichester, England: Wiley.

Renner, B. (2006). Curiosity about people: The development of a social curiosity measure in adults. *Journal of Personality Assessment, 87,* 305–316. http://dx.doi.org/10.1207/s15327752jpa8703_11

Robinson, J. D., & Heritage, J. (2006). Physicians' opening questions and patients' satisfaction. *Patient Education and Counseling, 60,* 279–285. http://dx.doi.org/10.1016/j.pec.2005.11.009

Rogers, C. R. (1951). *Client-centered therapy*. Boston, MA: Houghton Mifflin.

Rogers, C. R. (1955). *Active listening*. Chicago, IL: University of Chicago, Industrial Relations Center.

Rogers, C. R., & Farson, R. E. (2007). Active listening. In S. D. Ferguson & S. Ferguson (Eds.), *Organizational communication* (2nd ed., pp. 319–334). Piscataway, NJ: Transaction.

Rogers, T., & Norton, M. I. (2011). The artful dodger: Answering the wrong question the right way. *Journal of Experimental Psychology: Applied, 17,* 139–147. http://dx.doi.org/10.1037/a0023439

Rogers, T., Zeckhauser, R., Gino, F., Norton, M., & Schweitzer, M. (2016). (in press). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology*.

Rosenfeld, H. M. (1966). Approval-seeking and approval-inducing functions of verbal and nonverbal responses in the dyad. *Journal of Personality and Social Psychology, 4,* 597–605. http://dx.doi.org/10.1037/h0023996

Schegloff, E. A. (2007). *Sequence organization in interaction: Vol. 1. A primer in conversation analysis*. New York, NY: Cambridge University Press.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8,* 289–327. http://dx.doi.org/10.1515/semi.1973.8.4.289

Sedikides, C., Campbell, W. K., Reader, G. D., & Elliot, A. J. (1999). The relationship closeness induction task. *Representative Research in Social Psychology, 23,* 1–4.

Sezer, O., Gino, F., & Norton, M. I. (2015). *Humblebragging: A distinct—and ineffective—self-presentation strategy*. (Working Paper No. 15–080). Boston, MA: Harvard Business School.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

Silvia, P. J. (2008). Interest—The curious emotion. *Current Directions in Psychological Science, 17,* 57–60. http://dx.doi.org/10.1111/j.1467-8721.2008.00548.x

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

Sprecher, S. (1998). Insiders' perspectives on reasons for attraction to a close other. *Social Psychology Quarterly, 61,* 287–300. http://dx.doi.org/10.2307/2787031

Sprecher, S., Treger, S., Wondra, J. D., Hilaire, N., & Wallpe, K. (2013). Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology, 49,* 860–866. http://dx.doi.org/10.1016/j.jesp.2013.03.017

Stanley, S. M., Bradbury, T. N., & Markman, H. J. (2000). Structural flaws in the bridge from basic research on marriage to interventions for couples. *Journal of Marriage and the Family, 62,* 256–264. http://dx.doi.org/10.1111/j.1741-3737.2000.00256.x

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological), 36*, 111–147.

Surra, C. A., & Longstreth, M. (1990). Similarity of outcomes, interdependence, and conflict in dating relationships. *Journal of Personality and Social Psychology, 59,* 501–516. http://dx.doi.org/10.1037/0022-3514.59.3.501

Swann, W. B., Jr., & Gill, M. J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology, 73,* 747–757. http://dx.doi.org/10.1037/0022-3514.73.4.747

Swann, W. B., Jr., Stephenson, B., & Pittman, T. S. (1981). Curiosity and control: On the determinants of the search for social knowledge. *Journal of Personality and Social Psychology, 40,* 635–642. http://dx.doi.org/10.1037/0022-3514.40.4.635

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58,* 267–288.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R Package for causal mediation analysis. *Journal of Statistical Software, 59,* 1–38. http://dx.doi.org/10.18637/jss.v059.i05

Van der Meij, H. (1987). Assumptions of information-seeking questions. *Questioning Exchange, 1,* 111–117.

Vangelisti, A. L., Knapp, M. L., & Daly, J. A. (1990). Conversational narcissism. *Communication Monographs, 57,* 251–274. http://dx.doi.org/10.1080/03637759009376202

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics, 7,* 91. http://dx.doi.org/10.1186/1471-2105-7-91

Weger, H., Jr., Bell, G. R. C., Minei, E. M., & Robinson, M. C. (2014). The relative effectiveness of active listening in initial interactions. *International Journal of Listening, 28,* 13–31. http://dx.doi.org/10.1080/10904018.2013.813234

Wilson, W., & Henzlik, W. (1986). Reciprocity of liking following face-to-face encounters with attractive and unattractive others. *Psychological Reports, 59,* 599–609. http://dx.doi.org/10.2466/pr0.1986.59.2.599

Appendix A

Measures from Studies 1 and 2

Here, we report all measures that were collected during the studies. We discuss the findings from our main measures in the manuscript. For methodological transparency, we report other measures we collected, including those not connected to the main results.

Pairing Check (Studies 1, 2A)

Were you paired with another participant? Note: If there were technical issues with the chat, it will not affect your payment.

Yes, I was joined in the chat with another participant.
No, no one ever joined me in the chat.

Chat Type (Studies 1, 2A)

If you were paired with a participant, how would you best describe your interaction with the other participant? Note: Your responses will not affect your payment.

The other participant and I had a full conversation.
The other participant only said a few lines, then didn't respond.
I only said a few lines, then didn't respond.
The other participant only said a few lines, then left the chat.
The other participant never responded.
I never responded to the other participant.

Reading Check (Study 2B)

Were you able to read the conversation between User 1 and User 2? (Note: If you were not able to read the conversation, you will still be paid for participating in the study.) (Yes/No).

(*Appendices continue*)

The following interpersonal perception measures are rated on 1 to 7 scales unless noted otherwise.

Activities Preferences Questionnaire Self (Study 1)

Please, tell us how much you enjoy doing each of the following activities. That is, give us your own, personal opinion for each question.

1 (*dislike extremely*) through 9 (*like extremely*)

Reading
Watching TV
Partying
Seeing live music
Watching sports
Playing board games
Cooking
Working out
Cleaning

Activities Preferences Questionnaire Partner (Study 1)

We asked your partner to tell us how much they "enjoy doing each of the following activities." Now we'd like you to guess how your partner responded. That is, what answer did they just give to these exact questions?

(Same items as APQ Self)

Liking (Studies 1, 2A)

My partner is likable.
I liked my partner.
I would enjoy spending time with my partner.
I dislike my partner (reverse-scored).

Predicted Liking (Studies 1, 2A)

My partner thinks I'm likable.
My partner liked me.
My partner would enjoy spending time with me.
My partner dislikes me (reverse-scored).

Ratings of Liking (Study 2B)

Liking of User 1.

User 1 is likable.
I like User 1.
I would enjoy spending time with User 1.
I dislike User 1.

Liking of User 2.

User 2 is likable.
I like User 2.
I would enjoy spending time with User 2.
I dislike User 2.

Ratings of Predicted Liking (Study 2B)

Prediction of User 1's response.

User 1 thinks User 2 is likable.
User 1 liked User 2.
User 1 would enjoy spending time with User 2.
User 1 dislikes User 2.

Prediction of User 2's response.

User 2 thinks User 1 is likable.
User 2 liked User 1.
User 2 would enjoy spending time with User 1.
User 2 dislikes User 1.

Enjoyment (Study 1)

I enjoyed this conversation.
I thought this conversation was engaging.
I had an interesting conversation with this person.

Predicted enjoyment (Study 1).

My partner enjoyed this conversation.
My partner thought this conversation was engaging.
My partner had an interesting conversation with me.

Perspective-Taking (Study 2A)

I understand this person's situation.
I can imagine being in this person's place.
I can easily imagine how things look from this person's perspective.

Perceived perspective-taking (Study 2A).

This person understands my situation.
This person can imagine being in my place.
This person can easily imagine how things look from my perspective.

(*Appendices continue*)

Empathic Concern (Study 2A)

I feel warm toward this person.
I feel compassion for this person.
I feel sympathetic toward this person.

Perceived empathic concern (Study 2A).

This person feels warm toward me.
This person feels compassion for me.
This person feels sympathetic toward me.

Confidence (Study 1)

How well do you think you know what activities your partner would enjoy?

How well do you think your partner knows what activities you would enjoy?

Self–Other Similarity (Study 1)

How much do you think you have in common with your partner?

How similar do you think you and your partner are likely to be?

Estimated Questions (Studies 1, 2A)

During the conversation with your partner, how many questions did *you* ask? Please give your best estimate in the following box:

During the conversation with your partner, how many questions did *your partner* ask? Please give your best estimate in the following box:

Estimated Questions (Study 2B)

In the conversation you just read, how many questions did *User 1* ask? Please give your best estimate in the box below:

In the conversation you just read, how many questions did *User 2* ask? Please give your best estimate in the following box:

Manipulation Check (Studies 1, 2A)

Were you given instructions to ask questions? (Yes/No)

How many questions were you instructed to ask?

Questions wanted to ask (Studies 1, 2A)

Did you want to ask more questions or fewer questions than you were instructed to ask? (More/Fewer/About the same)

During the conversation with your partner, how many questions did you actually want to ask? Please enter a number in the box below:

Demographics (Studies 1, 2A, 2B)

What is your gender? (Male/Female)

What is your age (in years)?

Dictator Game (Study 2A)

You have been granted a $1.00 bonus (in addition to your $2.00 base pay for participating in this study).

You must now decide how much of the $1.00 to give to your partner, and how much to keep for yourself. You get to keep the amount that you do not give to your partner.

You can give some, all, or none of the $1.00 to your partner. Your partner will then keep the amount you give to them. (The payments are given after the survey is finished.)

Your partner will NOT know the amount of money you give until after the survey is finished. You will never see, meet, or interact with your partner in the future.

How much do you decide to give to your partner? Enter a number between $0.00 and $1.00 in the following box:

(*Appendices continue*)

Appendix B

Follow-Up Question Detection Algorithm

In this section, we report our development of the follow-up question detector. This algorithm serves two purposes. First, we can classify question types automatically, so that we do not need manual coding for large-scale data (as in Study 3). Second, we can learn what features of follow-up questions distinguish them from other questions, providing greater insight than holistic manual codes. These distinctive features inform our academic understanding, and also provide a practical, prescriptive guide for those who want to ask more follow-up questions. We apply this question-type detector to our data from Study 3, but others may apply this detector to any conversational data more broadly. Please contact the authors for more information.

We follow a series of standard text analysis procedures (Grimmer & Stewart, 2013; Jurafsky & Martin, 2009), which are reported here in detail. First, we construct a training set from the manually coded (human-coded) data in Studies 1 and 2. Second, we preprocess the question text in the training set to produce a high-dimensional feature space. Third, we use a machine learning algorithm to learn the distinctive features of each question type. Fourth, we apply the trained algorithm to out-of-sample data, to generate classifications for new questions. Finally, we estimate the accuracy of this procedure using nested cross-validation within the training set.

Training Set

We approached the detection algorithm as a supervised learning problem. Specifically, the data from Study 1 and Study 2 were taken as ground truth examples of the question types we had asked our coders to classify. Across both data sets, we extracted all conversational turns that included a question (n = 4,545), along with the hand-coded label as one of six types: introductory (n = 251), mirror (n = 865), full-switch (n = 1252), follow-up (n = 1,841), partial-switch (n = 249), or rhetorical (n = 87). The partial-switch and rhetorical types were too rare here to make any reliable inferences (and were not relevant to our hypotheses), so we decided to drop them.

The resulting training dataset of question turns (n = 4,209) was composed of two measures. The outcome measure was a four-class multinomial label: introductory, mirror, switch, or follow-up. The prediction measure was the contents of the text in each question turn. The goal of the algorithm, then, would be to take any new conversational turn that included a question and assign it one of the four labels from the training data. To accomplish this, we required an automated method for turning the unstructured text data into structured numeric data.

Text Processing

The text from each question turn was parsed according to the following steps, using the R package “tm” (Feinerer, Hornik, & Meyer, 2008). In order, the text was converted to lowercase; then contractions were expanded; then punctuation was removed (with the exception of question marks and exclamation points, which were treated as though they were words). The remaining words were stemmed using the standard Porter stemmer, and then grouped into “ngrams,” groups of one, two, or three sequential word stems. For example, “how are you doing?” would be parsed into 12 stemmed ngrams [“how”, “are”, “you”, “do”, “?”, “how are”, “are you”, “you do”, “do?” “how are you”, “are you do”, and “you do ?”].

One exception to the standard natural language processing workflow is that stop words were not removed, and this was important for several reasons. First, they form a disproportionate amount of the words used in conversation—in fact 42% of all ngrams in our training set were stop words, including all of the question words (“who,” “what,” “where,” “when,” “why,” “how,” and “which”). Furthermore, we wanted to learn the syntactic structure of questions, rather than extracting semantic content. In our data we observe that many prevalent phrases that form the structure of questions are composed entirely of common stop words. This would increase accuracy within a dataset and also make our results more robust across conversations in different contexts and people, where the semantic content might be entirely different.

The primary feature extraction method followed a “bag of words” approach, which simply counted the ngrams from each question, removing all information about the order in which the ngrams appear. To focus on the most prevalent features, ngrams which appeared in less than 1% of all questions were excluded. This process reduced the documents to a “feature count matrix,” in which each question turn was assigned a row, while each ngram feature was assigned a column, and the value of each cell represented the number of times that ngram appeared in that document. This dataset is sparse—specifically, 96% of the cells are zero, since most questions only include a few of the 372 ngrams that were extracted during this process.

In addition to the ngram counts, we calculated six holistic features for each question turn. This included the word count of the entire turn, as well as whether multiple questions were asked in that turn. We also measured the distance of the turn from the beginning of the conversation, as a fraction of the total conversation length. Finally, we included variables indicating whether the question was preceded by a statement in the same turn, whether the question-asker’s previous turn included a question, and the question target’s previous turn included a question.

Classifier Algorithm

These steps processed the unstructured text into a high-dimensional set of features. So we needed an algorithm that could automatically determine which subset of features were the most

(Appendices continue)

Table B1
Model Coefficients for Holistic Features

Holistic feature	Average	Follow-up	Full-switch	Introductory	Mirror
Word count of turn	16.94	.20	.00	.00	.00
Distance into the conversation	42.89%	.25	.00	−1.34	.00
Asker's last turn was question	49.16%	.00	.00	−.43	.61
Asker's last turn was question	54.00%	.30	.12	−.33	−.12
Multiple questions in turn	9.81%	.08	−.04	.00	.00
Pre-question statement in turn	32.79%	−.10	.00	.00	.37

useful for classifying question types. We use a common method, the LASSO, implemented in the *glmnet* package (Tibshirani, 1996; Friedman, Hastie, & Tibshirani, 2010). This algorithm estimates a multinomial logistic regression, and regularizes the effective feature space by imposing a constraint on the total absolute size of the coefficients across all features. The size of that constraint is determined empirically, by minimizing out-of-sample error via cross-validation within the training set. This process reduces many coefficients in the regression to exactly zero, leaving a smaller set with nonzero coefficients in the model.

Although the model is too complex to report in full, we provide two tables for edification. In Table B1 we report the coefficients that map the holistic features to each question type, along with the average for the feature. The results are intuitive—compared with other question types, follow-up questions are longer, asked later in the conversation, follow a previous question from the asker, and are somewhat more likely to contain multiple questions in the same turn, but are not likely to be preceded by a statement in that turn.

In Table B2, we report distinctive ngram features for each question type. Again, the results are intuitive. Turns with follow-up questions involve appreciation of the previous statement ("nice," "cool," "wow"), or question phrases that prompt elaborations ("why . . .," "which," "what kind . . .," "how do . . .," "where do . . .," "is it . . ."). Full-switch questions are open-ended, and focus on generic facts (e.g., "where [are you from/do you live]?", "what do you like . . .," "interests," "hobbies?"). Introductory questions (e.g., "Hi, how are you today?") and mirror questions (e.g., "Yes, i am great, and how about yourself?") are simpler categories, that follow conventional and distinctive scripts. Although the full model contains far more complexity than this table alone can provide, it is reassuring to know that the features selected by the model align with basic intuition.

Predictions

Once the model was built, any new test set of questions could be labeled according to this four-class scheme. To do so, the text in the test set would simply have to be processed in the same way as the text in the training set, tallying the presence of the same 372 ngrams in the new text (along with the six holistic features). Note that the prevalence filter was still applied with respect to the training set, not the test set—that is, only the 372 ngrams that were included in the training set were counted, no matter how rare or common they were in the test set.

Table B2
Distinctive Features for Each Question Type

Follow-up	Full-switch	Introductory	Mirror
Which	How old	How are you	How about
Why	Do you like	Hello	What about
What kind	Travel	Your name	Yourself?
Cool	Fun	How are	And
Nice	Do you live	Hi how	I am
Wow	Interests	Today?	And you
Is it	Hobbies?	What is	What about you
How do	You a student	Go?	And
Where do	Weather	Name?	No, I
Want to	You from?	Are you?	Yes, I

(*Appendices continue*)

Table B3
Confusion Matrix of Predicted and Actual Four-Class Question Types

Question type	True label of question			
	Follow-up	Full-switch	Introductory	Mirror
Predicted label of question				
Follow-up	1,377	358	7	99
Full-switch	373	806	30	139
Introductory	2	17	202	18
Mirror	89	71	12	609

For every new conversational turn, the algorithm provided probabilistic multinomial likelihoods across the four classes. As an example, the turn "I see. Why did you have to move?" was given the following likelihoods: introductory = 0%; mirror = 4%, switch = 20%, follow-up = 76%. In our regressions, we use these estimated probabilities as direct predictor variables, to account for the relative uncertainty the model might have about some question turns. However, we confirmed that the results of our analyses were identical when we used the probabilistic model output to apply a single deterministic label to each question turn, and then use those binary labels as predictors in a regression.

Accuracy Validation

Before applying the algorithm to new unlabeled data, we wanted to estimate the algorithm's accuracy on the labeled data in the training set. We did this using a nested cross-validation procedure (Stone, 1974; Varma & Simon, 2006). The entire dataset was randomly divided into 10 folds of equal size. To produce out-of-sample predictions for each fold, a classification model was trained and tuned on the other nine folds, and applied directly to the held-out data to estimate the likelihood that each held-out question was a follow-up question. To smooth out the random fluctuation, we performed this procedure five times, and averaged across all five likelihoods, to determine a final estimated likelihood for each question.

We measure the accuracy of this procedure in two ways. First, we consider its ability to classify follow-up questions as distinct from the three other question types (i.e., mirror, introductory, and full switch), which were lumped together. The resulting two-class likelihood (follow-up vs. other-type) was compared to the true binary labels (follow-up vs. other-type) for each question using the area under the curve metric (AUC). This test calculates the probability that any follow-up question will be assigned a higher follow-up likelihood than any non-follow-up question (with 50% as a baseline of random guessing). Across the entire training set, this validation exercise produced an accuracy level of 86.97% [95% CI: 85.92%, 88.01%], which we considered to be a high level of accuracy in this context.

As a second measure of model accuracy, we consider all four question types by assigning a deterministic label to each question type. However, we must account for the fact that the base rate of follow-up questions (44%) is nearly double that of any other question type. To do so, we use a two-tiered labeling process. First, we assign the "follow-up" label to the 44% of questions that are most likely to be follow-up questions, according to the model. For the remaining 56% of questions, we assign labels based on which of the remaining three labels is most likely, according to the model.

These predicted and true labels are compared in a "confusion matrix" in Table B3. The multinomial classification accuracy is given by the diagonal cells (71.13%), and the implied binary accuracy of classifying follow-up versus non-follow-up questions is higher still (77.95%). These results are somewhat lower than the binary accuracy above, as should be expected. But overall, these results suggest that the model does not have any strong biases toward any particular misclassifications, and gives us further confidence in our ability to classify question types.

Received February 29, 2016
Revision received March 23, 2017
Accepted March 23, 2017 ■