

In press, *Journal of Personality and Social Psychology*

Overly Shallow?:

Miscalibrated Expectations Create a Barrier to Deeper Conversation

Michael Kardas

Northwestern University

Amit Kumar

University of Texas at Austin

Nicholas Epley

University of Chicago

Abstract

People may want deep and meaningful relationships with others, but may also be reluctant to engage in the deep and meaningful conversations with strangers that could create those relationships. We hypothesized that people systematically underestimate how caring and interested distant strangers are in one's own intimate revelations and that these miscalibrated expectations create a psychological barrier to deeper conversations. As predicted, conversations between strangers felt less awkward, and created more connectedness and happiness, than the participants themselves expected (Experiments 1a-5). Participants were especially prone to overestimate how awkward deep conversations would be compared to shallow conversations (Experiments 2-5). Notably, they also felt more connected to deep conversation partners than shallow conversation partners after having both types of conversations (Experiments 6a-b). Systematic differences between expectations and experiences arose because participants expected others to care less about their disclosures in conversation than others actually did (Experiments 1a, 1b, 4a, 4b, 5, 6a). As a result, participants more accurately predicted the outcomes of their conversations when speaking with close friends, family, or partners whose care and interest is more clearly known (Experiment 5). Miscalibrated expectations about others matter because they guide decisions about which topics to discuss in conversation, such that more calibrated expectations encourage deeper conversation (Experiments 7a-7b). Misunderstanding others can encourage overly shallow interactions.

Keywords: self-disclosure, intimacy, social connection, accuracy, social cognition

Word Count: 213

Overly Shallow?:

Miscalibrated Expectations Create a Barrier to Deeper Conversation

“A number of porcupines huddled together for warmth on a cold day in winter; but, as they began to prick one another with their quills, they were obliged to disperse. However the cold drove them together again, when just the same thing happened . . . In the same way the need of society drives the human porcupines together, only to be mutually repelled by the many prickly and disagreeable qualities of their nature. The moderate distance which they at last discover to be the only tolerable condition of intercourse, is the code of politeness and fine manners.”

—Schopenhauer, 1851/1964, p. 226

Schopenhauer argued that people want deeper and more intimate relationships but are reluctant to pursue them because they expect that greater intimacy will be unpleasant. These expectations, in turn, discourage intimacy, entrenching the “politeness and fine manners” of small talk as the norm in everyday discourse. Indeed, everyday conversation seems to be anything but a stream of deep and meaningful exchanges: Fewer than half of people’s conversations are substantive and meaningful (Mehl, Vazire, Holleran, & Clark, 2010), creating an entire genre of self-help books promising to help you make better “small talk” and “chit chat” with those around you.

The fine manners of “small talk” aside, empirical research suggests notable benefits of having more “deep talk” in everyday life. Deep and meaningful conversations strengthen social ties (Aron, Melinat, Aron, Vallone, & Bator, 1997; Collins & Miller, 1994), relieve the psychological burdens of secrecy or negative emotional experiences (Pennebaker, 1997; Slepian

& Moulton-Tetlock, 2019), and speed the development of close relationships (Altman & Taylor, 1973; Derlega, Winstead, & Greene, 2001). Presumably because positive social relationships bring happiness and wellbeing (e.g., Diener & Seligman, 2002), those who spend more time engaging in “deep talk” tend to be happier than those who spend relatively more time in “small talk” (Mehl et al., 2010; Milek et al., 2018). With all of the positive outcomes that spring from deep talk, what keeps people from having more of it in their everyday lives?

Here we examine the psychological processes guiding conversations to understand why people choose to discuss relatively deep versus shallow topics, and whether those choices accurately anticipate the outcomes of relatively deep and shallow conversations. We define deep conversations as those in which two people engage in self-disclosure by revealing personally intimate information about their thoughts, feelings, or experiences. We predicted that people’s decisions to engage in deep conversation are guided by their expectations about how much their interaction partner cares about the intimate details of one’s life, and that people are therefore reluctant to engage in deep talk to the extent they believe their conversation partner will be indifferent to the conversation’s content. Perhaps more important, we also predict that people’s expectations are systematically miscalibrated, such that they underestimate how much others care about one’s own disclosures in conversation, and thereby overestimate how awkward and uncomfortable deep conversations will be. We suggest that these miscalibrated expectations about others can create a psychological barrier to engaging in deep conversations more often in everyday life.

Psychological Barriers to Meaningful Conversation

Belonging is a fundamental human need virtually on par with eating and sleeping (Baumeister & Leary, 1995; Ryan & Deci, 2000), and yet previous research suggests that people

often forego opportunities to connect with others that would satisfy this need and thereby enhance both their own and others' wellbeing. People can be somewhat reluctant to talk with strangers (Dunn, Biesanz, Human, & Finn, 2007; Epley & Schroeder, 2014; Mallett, Wilson, & Gilbert, 2008), give compliments (Boothby & Bohns, 2020; Zhao & Epley, in press), express gratitude (Kumar & Epley, 2018), perform random acts of kindness (Dunn, Aknin, & Norton, 2008; Kumar & Epley, 2021), or disclose personal information (John, Barasz, & Norton, 2016; Kardas, Kumar, & Epley, 2021), at least partly because people in these cases underestimate the positive impact their social behavior will have on others. Even the briefest engagements with others, from saying "hello" to a barista (Sandstrom & Dunn, 2014a; 2014b) to making eye contact with a passerby (Wesselman, Cardoso, Slater, & Williams, 2012), can increase the sense of connection to others, and yet people sometimes forego even these easy opportunities to engage with others. Misunderstanding how positively others value social interaction can create a barrier to engaging with others, thereby creating a systematic barrier to satisfying the basic human need for belonging.

Whereas existing research has primarily examined how wisely people choose *whether* to engage with others, we expand this body of work by examining how wisely people choose *how* to engage with others in the midst of conversation. Any conversation can range from relatively shallow and superficial to relatively deep and intimate. Like many decisions that are guided by some assessment of expected costs and benefits (Becker, 1993), we predict that people choose the depth of conversation based partly on how they expect that an intimate versus shallow conversation is likely to unfold.

Although there are likely to be multiple mechanisms that influence people's expectations about a conversation, we theorize that people's expectations about shallow and deep

conversation are guided by at least two inferences. First, expectations should be guided by the information people expect to share during the conversation. Shallow conversations tend to be impersonal and require relatively little self-disclosure. People are unlikely to feel vulnerable to others' evaluations while discussing impersonal topics like the weather. Deep conversations, in contrast, require sharing personal information about one's past experiences, preferences, or beliefs, which could leave people feeling more vulnerable to others' evaluations (Berenson et al., 2009; Leary, 1983). Existing research suggests that these fears of vulnerability may be misplaced. In one series of experiments, people overestimated how harshly they would be judged by others when they were in the midst of some embarrassing blunder or mishap (Savitsky, Epley, & Gilovich, 2001). In another series of experiments, people overestimated how negatively they would be judged when they revealed personal imperfections or weaknesses to another person (Bruk, Scholl, & Bless, 2018). Finally, strangers talking for the first time reported feeling more positive, less awkward, and more connected to each other after a relatively deep conversation than they expected, especially when they were communicating over a relatively more intimate (voice-based) communication media compared to less intimate (text-based) media (Kumar & Epley, 2021, Experiment 2). This research did not include a shallow conversation condition for comparison nor did it examine underlying mechanisms that might explain why people would undervalue deep conversation. Collectively, this existing research suggests that deeper conversations may not leave us as vulnerable as we might expect.

Second, expectations should also be guided at least in part by people's inferences about how much one's conversation partner will care about the intimate and meaningful topics that make up deeper conversation. Whereas a shallow conversation may be unlikely to feel awkward or uncomfortable regardless of how caring and interested the recipient is, a deep conversation

may feel quite awkward with an uncaring or indifferent partner who cares little for learning more about us. This predicts that people should have more positive expectations about deep conversation when they expect their conversation partner to be caring and interested than when they expect another person to be uncaring or indifferent. People should also be more willing to discuss deeper topics when they expect their partner to be relatively caring and interested in what one has to say.

Identifying how wisely people choose the depth of their conversation topics therefore requires understanding how accurately people infer both their own and others' care and interest in the content of deep conversation. People can directly assess, via introspection, how much they personally care about what others have to say regarding deep and intimate topics, but have to infer how much others care (Epley & Waytz, 2010; Gilbert & Malone, 1995; Jones, 1979; Nisbett & Wilson, 1977). People are also deeply motivated to connect with others (Baumeister & Leary, 1995), and deeply value signs of trustworthiness and honesty from others, including others' willingness to open up in deeper conversation and their willingness to listen attentively to one's own statements (Ames, Maissen, & Brockner, 2012; Cuddy, Fiske, & Glick, 2008; Kluger et al., 2020; Kluger & Zaidel, 2013; Levine & Cohen, 2018). One series of experiments even indicated that people underestimate how much attention others pay to them in everyday life (Boothby, Clark, & Bargh, 2017), suggesting that people may fail to fully appreciate how much interest others are likely to take in them.

This prior research led us to make several predictions about the accuracy of people's expectations about care and interest. First, we predicted that people would assume that they would care more about the intimate details of what they share in the midst of a deep conversation than a typical stranger would care about those same intimate revelations. Second, we predicted

that people's beliefs about their own interest and care would be more calibrated than their inferences about a stranger's interest, thereby leading people to underestimate others' interest and care (but not their own). Third, we predicted that underestimating the extent to which others are interested in, and care about, one's statements in conversation would also lead people to underestimate how connected they would feel to their conversation partner and how happy they would feel about the conversation, and to overestimate how negative the conversations would be. Fourth, because conversations with seemingly indifferent others could feel awkward when they require more intimate self-disclosure, we predicted that participants would overestimate the awkwardness of deep conversation *more* than the awkwardness of shallow conversation. Finally, because one's friends are known to care more about oneself than strangers, we expected that people would be more calibrated predicting the outcomes of their conversations with friends than with strangers because people would anticipate more positive outcomes with friends, even when discussing relatively intimate topics.

We believe that misunderstanding others' sociality—the degree to which others are interested in, and care about, connecting through conversation—matters because it creates a psychological barrier to engaging in deeper and more meaningful conversation. This predicts that people are overly reluctant to engage in deep conversation because they have miscalibrated expectations about the consequences of sharing meaningful information with others whose interest and care is unknown. A simple alternative interpretation is that people don't engage in deeper conversations with strangers because they do not care about what the *other* person has to say. If people are indifferent to strangers' responses in deep conversations, and hence are uninterested in having deeper conversations with them, then none of the hypotheses we

articulated above would be worth testing because beliefs about others' care would not guide people's decisions to engage in relatively shallow versus deep conversation.

We therefore conducted an initial pilot experiment using Amazon's Mechanical Turk ($N = 188$) in which participants imagined talking with someone they had never met before and were asked how interested they were in sharing intimate details with a stranger or learning those same details about a stranger. Specifically, we randomly assigned participants to one of four roles: *Listeners* imagined that they would listen to the other person's responses to a series of questions; *Askers* imagined that they would ask the other person a series of questions and then listen to the person's responses; *Discussants* imagined that they would both ask and answer a series of questions with the other person; and *Answerers* imagined that they would answer a series of questions while the other person listened. Participants then viewed a list of 20 pre-tested discussion questions that varied in intimacy from very shallow and superficial to very deep and intimate (see Appendix), and selected the 5 questions they preferred for the exchange. We hypothesized that Discussants and Answerers, who were required to answer the questions and reveal information about themselves, would choose less intimate questions on average than Listeners and Askers, who were simply able to learn about others.

Results supported these hypotheses. Listeners ($M_{\text{intimacy}} = 5.06$) and Askers ($M_{\text{intimacy}} = 4.71$) chose questions higher in average intimacy than either Discussants ($M_{\text{intimacy}} = 4.04$) or Answerers ($M_{\text{intimacy}} = 3.58$; $p \leq .008$). Whereas the majority of the questions selected by Listeners ($M = 3.06$ out of 5; $p = .017$) were "deep" questions that were more intimate than the median question, and Askers ($M = 2.74$; $p = .320$) selected equal numbers of "deep" and "shallow" questions, Discussants ($M = 1.89$; $p = .003$) and Answerers ($M = 1.36$; $p < .001$) chose fewer deep questions than shallow questions (see Supplemental Material for the full method and

results). This suggests that people are indeed interested in knowing intimate information about others (see also Hart, VanEpps, & Schweitzer, 2021), but are reluctant to reveal intimate information about themselves. A follow-up experiment ($N = 144$) suggested that this difference did not stem from differences in the perceived difficulty of generating answers to shallow versus deep questions, as participants' choices did not differ between a condition in which they imagined revealing pre-written responses ($M_{\text{intimacy}} = 3.79$) and one in which they imagined generating answers to the questions in real time ($M_{\text{intimacy}} = 3.83$; $p = .885$). Again, participants in the conditions that required revealing personal information in conversation preferred to discuss less intimate questions than participants who imagined only listening to the other person's responses ($M_{\text{intimacy}} = 4.80$; $ps < .001$; see Supplemental Material). These results indicate that these participants were more interested in getting to know meaningful information about others than they were in revealing meaningful information to others, suggesting a barrier to engaging in deep conversation that may come from inferences about how others will respond to these self-disclosures. This paper reports a series of experiments testing whether a reluctance to engage in deep conversation arises partly from miscalibrated concerns that strangers will be indifferent toward one's self-disclosures, creating more superficial conversations than might be ideal for one's own and others' wellbeing.

Of course, conversations can range from very shallow and superficial to very deep and intimate. Our hypotheses do not suggest that underestimating others' care creates a barrier to having the *deepest* conversations. Some topics may be sufficiently deep that people will prefer to avoid discussing them even with close others who they expect to be highly caring and considerate. Rather, our hypotheses suggest that underestimating others' care creates a barrier to having *deeper* conversations. By "deeper" conversations we mean those that are more intimate

than people's typical conversations and are instead closer to the depth of conversation that people report wishing they would experience more often. We will test each component of our reasoning empirically: whether people's typical conversations are less intimate than they wish they would be, whether people underestimate others' care during conversation, and whether experimentally manipulating people's beliefs about others' care influences the depth of conversation that they prefer.

Overview of Experiments

We conducted a series of experiments to test our hypotheses, and in doing so we recruited diverse samples of participants including primarily American undergraduate and master's degree students (Experiments 2, 3, 4a, 4b, 6a, 6b), community members (Experiments 3, 5), financial services employees and executives (Experiments 1a, 1b), international MBA students from around the world (Experiment 1c), and online participants (Experiments 7a, 7b). We first tested whether people underestimate the positivity of deep and intimate conversations with strangers by asking participants in Experiments 1a-1c to discuss a series of deep and intimate questions with a stranger. Participants reported how they expected to feel after these conversations, and we then compared participants' expectations against their actual experiences. Experiments 2-6b compared shallow conversations against deeper conversations by manipulating the intimacy of participants' conversations. Experiment 5 also compared conversations between strangers versus known family or friends to examine whether participants are especially likely to misunderstand the outcomes of deep conversations with strangers. Experiments 6a and 6b assigned participants to engage in *both* shallow and deep conversations as a more direct test of whether relatively deeper conversations build stronger connections between strangers than shallower ones. Finally,

Experiments 7a and 7b examined the extent to which expectations of others' interest and care create a psychological barrier to deeper conversations.

Our experiments make several novel contributions to our understanding of social cognition and interpersonal relationships. First, previous experiments test whether people appreciate the benefits of distant social connections with strangers (Dunn, Biesanz, Human, & Finn, 2007; Epley & Schroeder, 2014; Mallett, Wilson, & Gilbert, 2008) but do not measure or manipulate the intimacy of these conversations. Strangers tend not to engage in particularly intimate conversations with each other (Taylor, 1968), meaning that our experiments are an even stronger test of the potential benefits of distant social connections (Sandstrom & Dunn, 2014b). Second, a substantial body of literature examines how people make inferences about others' warmth (e.g., Cislak & Wojciszke, 2008; Fiske, Cuddy, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002; Russell & Fiske, 2008; Todorov, Pakrashi, & Oosterhof, 2009) but very little work examines the accuracy of these inferences. Our experiments test whether people systematically underestimate others' care and concern in the context of deep and meaningful conversations. Third, whereas ample research indicates that inferences about others' warmth impact *whether* people choose to engage with one another (e.g., Adolphs, Tranel, & Damasio, 1998; Cuddy, Fiske, & Glick, 2007; Todorov, 2008; Wojciszke, Bazinska, & Jaworski, 1998; see also Epley & Schroeder, 2014), our experiments enrich this literature by testing whether a related set of inferences influences *how deeply* people choose to engage. Our findings therefore open new avenues for research about contexts that may remove psychological barriers to engaging deeply. Finally, whereas some experiments suggest that deep and intimate conversations can quickly foster experiences of connectedness between strangers (e.g., Aron et al., 1997; Fishman & Gardner, 2017), our experiments move substantially beyond prior research by trying to

understand why people might be reluctant to engage in these conversations in the first place. Practically speaking, our research may offer useful guidance for people trying to behave more wisely in order to enhance their own wellbeing and strengthen their social relationships. Everyday conversations, we suggest, could be more superficial than would be optimal for one's own wellbeing.

To maximize statistical power, we recruited at least 50 participants or pairs per condition (Experiments 3, 5, 7a, 7b), or at least 30 groups of four participants per condition (Experiments 6a, 6b), for experiments conducted in laboratory, field, and online settings, and recruited as many participants as we could in experiments conducted during presentations (Experiments 1a, 1b, 1c, 2, 4a, 4b). To test the robustness of our effects, we also conducted multiple replication experiments (Experiments 1a-c; 4a and 4b). We report all measures and manipulations throughout the manuscript, and we report all analyses without exclusions in the Supplemental Material. All experiments were approved by the university's Institutional Review Board and we obtained informed consent from all participants. The Supplemental Material, surveys, data, and analysis script, as well as pre-registrations for Experiments 1b, 2, 3, 4a, 4b, 5, 6a, 6b, 7a, 7b, and several supplemental experiments, can be accessed at tinyurl.com/overly-shallow-osf.

Experiments 1a-1c: Deep Conversations

We first tested whether people consistently underestimate how positively their deep conversations with strangers would unfold. In Experiments 1a, 1b, and 1c, we controlled the content of the conversations by providing deep and intimate discussion questions. We hypothesized that deep conversations between strangers would feel less awkward, and would lead to stronger bonds, greater liking, and greater happiness, than people anticipated. Although these experiments do not enable direct comparisons between shallow and deep conversations,

they serve as initial tests of whether highly intimate conversations are more positive than strangers expect them to be. We test both shallow and deep conversations in the experiments that follow.

Furthermore, we began testing why people may misunderstand the outcomes of deep conversations. We predicted that strangers would underestimate how interested the other person would be during the conversation, and that this misunderstanding would help to explain why deep and intimate conversations unfold more positively than people expect.

We conducted these three experiments on meaningfully different samples of varying sizes in order to test the robustness and reliability of any experimental results. Experiment 1a included financial executives at a conference, 1b included managers and employees at a large financial services firm, and 1c included international MBA students participating from around the world in an online session. In all cases, the sample size was determined by the number of people who attended the session that comprised the experiment. Because of the similarity in both the procedures and results, we present the methods and results in detail only for Experiment 1a and report summaries of Experiments 1b and 1c in the main text. Full details of Experiments 1b and 1c are reported in the Supplemental Materials.

Method (Experiment 1a)

Participants. Twenty-five pairs of financial services executives were recruited during a session at a management conference ($N = 50$ individuals; $M_{\text{age}} = 48.92$; $SD_{\text{age}} = 7.55$; 18.00% female; 86.00% Caucasian). The executives entered the session unaware that they would be participating in any experiment.

Procedure. Participants opened a survey on their computers or mobile devices. They read that they would be randomly paired with another person attending the session who they

hadn't met before and would answer and discuss four questions with the other person.

Participants then read four deep questions, adapted from the "fast friends" paradigm (Aron et al., 1997):

1. For what in your life do you feel most grateful? Tell the other participant about it.
2. If a crystal ball could tell you the truth about yourself, your life, your future, or anything else, what would you want to know?
3. If you were going to become a close friend with the other participant, please share what would be important for him or her to know.
4. Can you describe a time you cried in front of another person?

After reading the questions, participants reported their expectations of the upcoming interaction. First, participants reported how interested they would be in hearing the other person's answers (0 = *not at all interested*; 10 = *very interested*) and also how interested they expected the other person would be in hearing the participant's own answers (0 = *not at all interested*; 10 = *very interested*). Participants then predicted how awkward they would feel during the conversation (0 = *not at all awkward*; 10 = *very awkward*), how strong a bond they would feel with the other person (0 = *weak, like a stranger*; 10 = *strong, like a new friend*), how much they would like the other person (0 = *not at all*; 10 = *very much*), and how happy they would feel about the conversation with the other person (0 = *not at all happy*; 10 = *extremely happy*).

Research assistants then randomly paired participants together with someone they did not know in the session (most were unfamiliar with each other), and gave each participant a card containing their pair's ID number and the four discussion questions. Participants then had a conversation for approximately 10 minutes and returned to their original seats to fill out another

survey asking them to report their actual experiences in the conversation. First, participants reported their own experiences using the same measures on which they reported their expectations before the conversation. They then reported their perceptions of their partner's experiences on these measures. Participants then thought about times when they had spoken with strangers and reported how often they wished they engaged in small talk with strangers, and how often they wished they engaged in deep conversations with strangers, on separate scales ranging from -5 (*Much less often than I do now*) to 0 (*Neither less nor more often than I do now*) to +5 (*Much more often than I do now*). Finally, participants reported demographic information and were debriefed.

Results (Experiment 1a)

The strength-of-bond and liking items were highly correlated in both expectations ($\alpha = .82$) and experiences ($\alpha = .85$) and so we combined these items to form a connectedness scale. In this and the following experiments, we computed mean responses across paired participants for each of the primary measures and then performed analyses at the level of the dyad.

Participants underestimated both their own interest in hearing from the other person, *paired* $t(24) = -5.92, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.29, -1.11], d = -1.36$, and how interested they would perceive their partner to be in hearing from them, *paired* $t(24) = -9.63, p < .001, 95\% \text{ CI}_{\text{difference}} = [-3.38, -2.18], d = -2.14$. Moreover, and consistent with our hypotheses, participants underestimated the extent to which their partners would be interested the content of the conversation significantly more than they underestimated their own interest, $F(1, 24) = 14.63, p < .001, \eta_p^2 = .38$.¹

¹ For all experiments, we present critical hypothesis tests in the main text and report all other main effects and interaction effects from ANOVAs in the Supplemental Material.

Our data suggest that these differences between participants' expectations and their own experiences reflect miscalibrated beliefs about a deep conversation rather than response biases that might arise in ratings of their partner's interest after the conversation. Specifically, participants' expectations of their partner's interest significantly underestimated both their partner's expected interest before the conversation, *paired t*(24) = -8.68, *p* < .001, 95% CI_{difference} = [-2.33, -1.43], *d* = -1.36, as well as their partner's self-reported interest after the conversation, *paired t*(24) = -14.07, *p* < .001, 95% CI_{difference} = [-4.11, -3.05], *d* = -3.04 (see Supplemental Material for further analyses).

Participants' conversations also unfolded significantly better than they anticipated: Participants overestimated how awkward they would feel, *paired t*(24) = 8.71, *p* < .001, 95% CI_{difference} = [2.44, 3.96], *d* = 1.74, underestimated how connected they would feel to their partner, *paired t*(24) = -11.32, *p* < .001, 95% CI_{difference} = [-2.34, -1.62], *d* = -2.26, and underestimated how happy they would feel about the conversation, *paired t*(24) = -9.59, *p* < .001, 95% CI_{difference} = [-2.99, -1.93], *d* = -1.92 (see Figure 1, top panel).^{2,3}

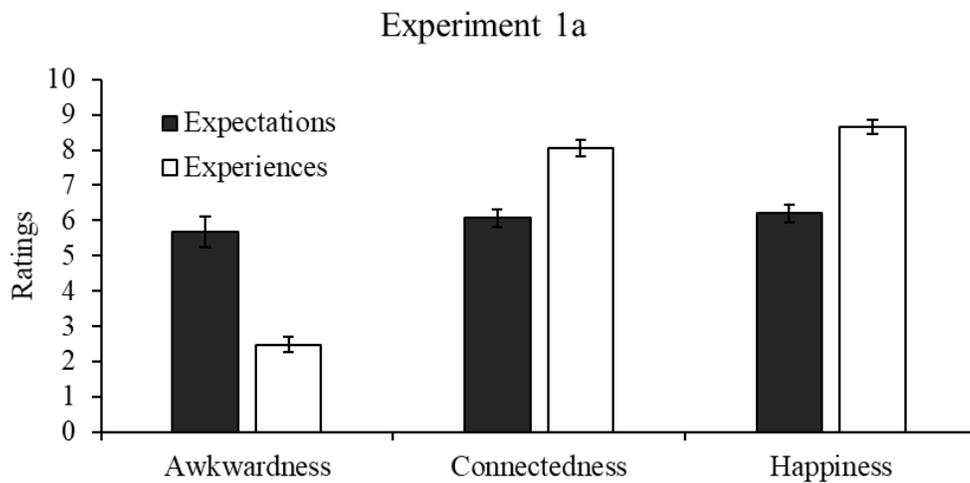
We hypothesized that underestimating the other person's interest would explain miscalibration on the primary measures. To test this, we performed within-pairs mediational analyses using measurement phase (expectation vs. experience) as the independent variable, partner interest as the mediating variable, and each of the primary measures as dependent variables in separate analyses using the MEMORE macro (Montoya & Hayes, 2017). The

² We conducted exploratory analyses using the actor-partner interdependence model (Kenny, Kashy, & Cook, 2006) to test whether miscalibration between participants' expectations and experiences was more strongly associated with the participants' own expectations (as our theory would suggest) or their partners' expectations. Throughout our experiments, we found consistent evidence that miscalibration was associated with one's own expectations, but little evidence that miscalibration was associated with one's partner's expectations (see Supplemental Table S2).

³ We also performed exploratory analyses of gender. Throughout our experiments, neither miscalibration on the primary measures nor differences in miscalibration between shallower and deeper conversations varied reliably between same-gender and mixed-gender pairs (see Supplemental Material for details).

indirect effects were significant for awkwardness ($b = -1.61$, 95% CI = [-3.08, -0.04]) and connectedness ($b = -0.79$, 95% CI = [-1.58, -0.13]) but non-significant for happiness ($b = -0.80$, 95% CI = [-1.97, 0.33]).

We also hypothesized that people want to engage in deeper conversations than they typically do. Consistent with this hypothesis, participants wished they engaged in marginally more small talk with strangers than they typically do ($M = 0.68$, $SD = 2.72$), $t(49) = 1.77$, $p = .083$, 95% CI = [-0.09, 1.45], $d = 0.25$, but wished they engaged in significantly more deep conversation with strangers than they typically do ($M = 2.60$, $SD = 1.93$), $t(49) = 9.54$, $p < .001$, 95% CI = [2.05, 3.15], $d = 1.35$. Moreover, participants' desire for more deep talk was significantly stronger than their desire for more small talk, *paired* $t(49) = 3.88$, $p < .001$, 95% CI_{difference} = [0.93, 2.91], $d = 0.55$. People may refrain from having more deep conversations in part because they underestimate how well these conversations will unfold.



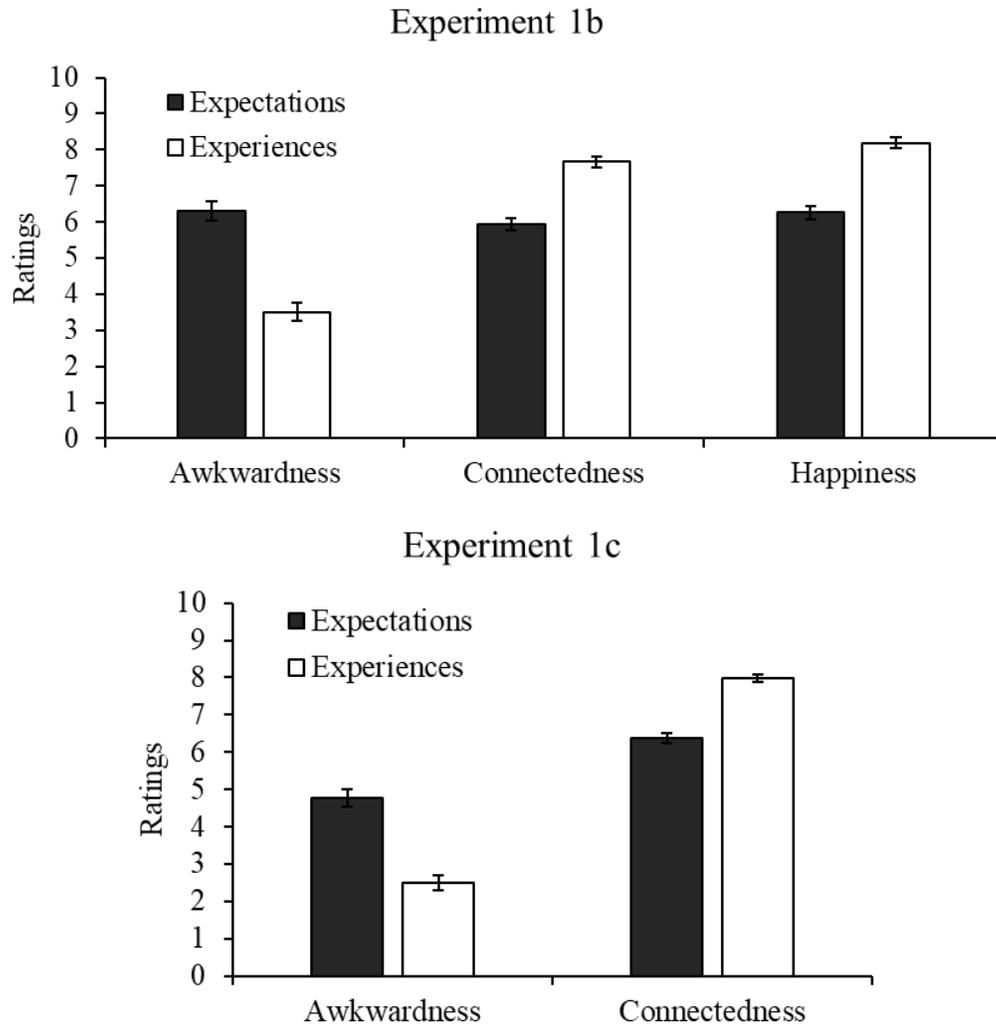


Figure 1. Mean awkwardness, connectedness, and happiness across measurement phase (expectations vs. experiences) in Experiments 1a-1c. Error bars $\pm 1 SE$.

Method & Results (Experiments 1b & 1c)

We also replicated these findings in Experiments 1b ($N = 50$ pairs) and 1c ($N = 56$ groups of 2-3 individuals; see Supplemental Material for the full method and results): Participants overestimated how awkward they would feel discussing deep topics with a stranger yet underestimated how connected they would feel to the other person and how happy they would feel about the conversation (see Figure 1). Experiment 1b also included measures of the other person's care and interest. In this experiment, participants significantly underestimated the

degree to which their partner would care about, and be interested in, the content of the conversation. The extent to which they did so statistically mediated the extent to which they underestimated their own feelings of connectedness and happiness in a deep conversation.

Experiment 2: Shallow Versus Deep Conversations

Experiments 1a-c suggest that relatively deep conversations between strangers tend to go better than expected. Specifically, participants felt less awkward, and felt happier and more connected to their partner, than they anticipated. These results were at least partly explained by participants' tendency to underestimate how caring and interested their partner would be in hearing what they had to say.

We suggest that the tendency to overestimate the conversation's awkwardness should be especially strong in deep conversations in which intimate self-disclosure is required, compared to relatively shallow conversations in which less intimate self-disclosure is required. We tested this hypothesis in Experiments 2-6b by manipulating the intimacy of conversation. In Experiment 2, we provided participants with a series of shallow or deep discussion questions, thereby manipulating the content of the conversation between pairs. We hypothesized that participants in the "deep talk" condition would overestimate how awkward their conversations would be more than participants in the "shallow talk" condition, and that participants in both conditions would underestimate how connected they would feel to their conversation partner.

Method

Participants. Eighty-nine pairs of master's degree students participated during a faculty-member presentation as part of a wellness event on campus ($N = 178$ individuals after exclusions; $M_{\text{age}} = 28.23$; $SD_{\text{age}} = 1.99$; 54.49% female; 53.37% Caucasian). Participants were unaware that they would be participating in an experiment until it was announced during the

presentation. An additional 7 participants could not be matched with their partner after the session because either 1 or 3 participants reported the same pair ID, and so were excluded from analyses. Because we analyzed the data at the level of the dyad rather than at the level of the individual, we did not re-analyze data with these participants included.

Procedure. Participants opened a survey on their computers or mobile devices and were randomly assigned to condition (signified by A or B) based on where they were seated in the room. Participants in Group A read the following five shallow questions adapted from Aron et al. (1997):

1. When was the last time you walked for more than an hour? Describe where you went and what you saw.
2. How did you celebrate last Halloween?
3. Do you like to get up early or stay up late? Why?
4. What is the best TV show you've seen in the last month? Tell your partner about it.
5. How often do you get your hair cut? Where do you go? Have you ever had a really bad hair cut experience?

Participants in Group B read the following five deep questions:

1. For what in your life do you feel most grateful? Tell your partner about it.
2. If a crystal ball could tell you the truth about yourself, your life, your future, or anything else, what would you want to know?
3. If you were going to become a close friend with your partner, please share what would be important for him or her to know.
4. What is one of the more embarrassing moments in your life?

5. Can you describe a time you cried in front of another person?

After reading the questions, participants predicted how awkward they would feel during the conversation, how uncomfortable, how much they would enjoy the conversation, how strong a bond they would feel with their conversation partner, how much they would like their conversation partner, and how well they would feel they got to know their conversation partner's "true beliefs, attitudes, preferences, and character," on similar 11-point scales as in Experiments 1a-1c. Participants were then randomly paired with someone from the same condition (shallow or deep) who they had not met before. Participants received question cards containing the five questions and were left to discuss the questions in pairs for about 10 minutes.

Participants then returned to their original seats and reported how they felt during their conversations on the same measures. To test whether deep conversation reduces momentary feelings of loneliness, participants then completed the 3-item perceived loneliness scale (Hughes, Waite, Hawkley, & Cacioppo, 2004), reporting (i) how often they feel they lack companionship, (ii) how often they feel left out, and (iii) how often they feel isolated from others (*Hardly ever* vs. *Some of the time* vs. *Often*). Because research has indicated that extraverts have better-calibrated beliefs about the benefits of extraverted social behavior than introverts (Zelenski et al., 2013), we then asked participants to complete the ten-item personality inventory (Gosling, Rentfrow, & Swann, 2003) in order to test whether calibration varies by personality. Finally, participants reported demographic information and were thanked and debriefed.

Results

The awkwardness and discomfort items were highly correlated ($\alpha_{\text{expectations}} = .98$, $\alpha_{\text{experiences}} = .77$), as were the enjoyment, strength of bond, liking, and perceived understanding

items ($\alpha_{\text{expectations}} = .69$, $\alpha_{\text{experiences}} = .87$). We therefore combined these items to form awkwardness and connectedness scales, respectively. We then conducted 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVAs with repeated measures on the second factor and awkwardness and connectedness as dependent measures in separate analyses.

Awkwardness. Participants in the deep condition overestimated how awkward their conversation would feel more than did participants in the shallow condition, as indicated by a significant conversation \times measurement phase interaction effect, $F(1, 87) = 71.34$, $p < .001$, $\eta_p^2 = .45$ (see Figure 2). Participants in the shallow condition expected their conversations to feel less awkward than participants in the deep condition, $t(87) = 11.58$, $p < .001$, 95% $CI_{\text{difference}} = [3.17, 4.48]$, $d = 2.49$, and also experienced less awkwardness, $t(87) = 5.22$, $p < .001$, 95% $CI_{\text{difference}} = [0.84, 1.87]$, $d = 1.12$. As predicted, the significant interaction indicates that participants in the deep condition ($d = 2.23$) overestimated how awkward their conversation would be more than did participants in the shallow condition ($d = 0.55$).

Connectedness. Participants in both the shallow and deep conditions underestimated how connected they would feel to their partner, as indicated by a significant main effect of measurement phase, $F(1, 87) = 98.24$, $p < .001$, $\eta_p^2 = .53$. The conversation \times measurement phase interaction effect was non-significant, $F(1, 87) = 0.12$, $p = .734$, $\eta_p^2 = .001$, indicating that participants underestimated how connected they would feel to their partner to similar degrees in the deep and shallow conditions.

Secondary analyses. We also tested whether participants in the deep condition felt less lonely overall at the end of their conversations than participants in the shallow condition by coding the response options on the perceived loneliness scale (*Hardly ever* vs. *Some of the time*

vs. *Often*) numerically from 1 to 3. Unexpectedly, participants in the deep condition ($M = 1.82$, $SD = 0.35$) reported feeling significantly more lonely than participants in the shallow condition ($M = 1.65$, $SD = 0.35$), $t(87) = 2.25$, $p = .027$, 95% $CI_{\text{difference}} = [0.02, 0.32]$, $d = 0.48$. We discuss this finding further in the study discussion below.

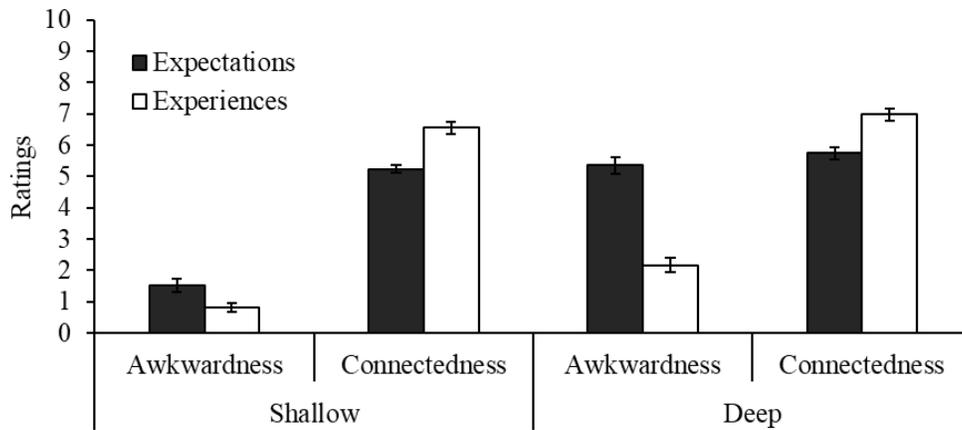


Figure 2. Mean awkwardness and connectedness across conversation conditions (shallow vs. deep) and measurement phase (expectations vs. experiences) in Experiment 2. Error bars $\pm 1 SE$.

We also tested how our results might be moderated by personality by regressing expected versus experienced awkwardness and connection simultaneously over the Big Five personality traits, for the shallow and deep conditions separately. We did not observe consistent moderation: The tendency to overestimate the awkwardness of a conversation was associated with higher neuroticism ($\beta = -0.26$, $t(98) = -2.58$, $p = .011$) among participants in the shallow condition but was not significantly associated with any Big Five traits among participants in the deep condition. Underestimating the sense of connection was associated with higher neuroticism ($\beta = 0.22$, $t(98) = 2.19$, $p = .031$) and lower openness ($\beta = -0.21$, $t(98) = -2.16$, $p = .033$) among participants in the shallow condition but was not associated with any Big Five traits among participants in the deep condition. Our findings were not moderated by extraversion for either

awkwardness or connectedness (see Supplemental Material for the full analyses). Personality did not consistently moderate the gap between expected and actual experience for either shallow or deep conversations.

Discussion

The results of this experiment replicate and extend the findings from Experiments 1a-1c: Participants' conversations felt less awkward, and led to greater feelings of connectedness, than the participants expected before engaging in the conversation. Furthermore, participants who engaged in deep conversations overestimated how awkward the conversation would be significantly *more* than those who engaged in shallow conversations. Although deep conversations did indeed feel more awkward than shallow conversations, the difference in participants' expectations was roughly four times larger than the difference that which participants actually experienced.

One other finding warrants further discussion: In this experiment, participants in the deep condition reported feeling significantly lonelier than participants in the shallow condition after their conversations. We did not predict this result, but we are also reluctant to conclude that this effect is a general consequence of deep conversations rather than an idiosyncratic consequence of the specific questions that participants discussed in this experiment. In particular, two of the deep questions, "What is one of the more embarrassing moments in your life?" and "Can you describe a time you cried in front of another person?", likely brought to mind memories of negative social experiences. These questions alone may have driven differences in reported loneliness. Because we are primarily interested in how connected people feel to their conversation partner, rather than more general feelings of loneliness, we did not include this measure in the following experiments but instead included measures that are more closely tied to

experienced wellbeing from the conversation itself. Specifically, we measured participants' feelings of happiness in Experiments 3-6. We also test a much wider range of questions in Experiments 3 and 5 to examine whether the effects observed in Experiment 2 generalize beyond the specific set of topics we asked participants to discuss.

Experiment 3: *Deeper Conversations*

Experiments 1a-2 indicated that deep conversations between strangers unfolded better than expected, and that participants are more likely to overestimate the awkwardness of deep conversations than shallow conversations. These results suggest that people might have what they perceive to be better conversations if they were willing to engage in deeper and more meaningful conversations in daily life. However, the preceding experiments do not directly test this potentially important practical conclusion for two reasons. First, we provided the conversation questions for participants. People may have more accurate beliefs about their conversations in daily life because they choose which topics to discuss. People may also be unable to generate deeper conversation topics on their own that would enhance their connections and wellbeing. Second, the preceding experiments lack a control condition to assess whether engaging in deeper conversations than one would normally engage in leads to better outcomes than expected. Assessing this is critical for understanding whether miscalibrated expectations could lead people to be overly shallow in conversation for their own wellbeing.

To address these limitations, participants in Experiment 3 wrote the questions that they later discussed with a stranger themselves. Participants first wrote a series of questions that they would normally discuss while getting to know someone new, and then wrote a series of questions that were deeper, involving topics that were more intimate than they would normally discuss. We then randomly assigned pairs to discuss the “control” questions or the “deeper” questions that

one of the participants had generated. As in Experiments 1a-2, we predicted that participants would underestimate how positive these conversations would leave them feeling, and that participants in the deep condition would overestimate how awkward the conversation would feel compared to participants in the control condition. This experiment also enables us to test whether generating the topics for discussion oneself, versus discussing topics that another person generated, affects the expected or actual experience of relatively shallow or deep conversations.

Method

Participants. We targeted 100 pairs of participants and finished recruiting once that target was reached after data exclusions. We achieved this by recruiting 103 pairs of participants from separate university and community subject pools ($N = 200$ individuals after exclusions; $M_{\text{age}} = 28.46$, $SD_{\text{age}} = 13.67$, 49.00% female, 31.50% Caucasian) to complete the experiment in exchange for \$6. We excluded 3 of these pairs from analyses because one pair knew one another beforehand, because one pair began their conversation before one member had reported expectations, and because one participant did not write out conversation questions.

Procedure. Participants sat in separate rooms and did not interact with one another prior to their conversations. Participants were told that they would develop questions that they might later ask and answer during a discussion. Both participants first generated five *control* questions. Specifically, they were told:

We would like you to begin by generating five questions. These should be the types of questions that you would naturally ask another person while first getting to know him or her. Please select questions that you would actually be willing to ask and answer later in this study, and these questions should be the types of questions that you would typically ask while first getting to know somebody.

The same participants then generated five *deep* questions. They were told:

Next we would like you to generate five more questions. This time, please generate five questions that are deeper and more intimate than the types of questions that you would naturally ask another person while first getting to know him or her. In other words, we would like you to generate questions that go beyond the surface, beyond small talk, to probe deeper subject matter that might be more personal or emotional. For example, you might ask the person about important experiences they've had or activities they've enjoyed. You might ask the person to reveal something important about them. These questions should require both you and your partner to reveal something about yourselves that you might not normally reveal in a conversation with a stranger. These should be topics that you would be more likely, perhaps, to talk about with a close friend or family member, and they should dig deeper than the ones you wrote down in the previous set of questions.

Pairs were then randomly assigned to either the control condition or the deep condition. In the control condition, the experimenter selected one of the two sets of control questions at random, whereas in the deep condition, the experimenter selected one of the two sets of deep questions at random. The participant who wrote the randomly selected questions then sequenced them in the order they preferred for the conversation.

Both participants then viewed the final set of five discussion questions on a computer screen. The participant who did not write the final set of questions knew that the questions were written by the other participant but was not told whether they were viewing the control or deep

questions. Participants then indicated how awkward they expected to feel during the conversation, how uncomfortable, how much they expected to enjoy the conversation, how strong a bond they would feel with their conversation partner, how much they would like their conversation partner, how well they would feel they got to know their conversation partner's true beliefs, attitudes, preferences, and character, and how happy they would feel about their conversation on 11-point scales as in the prior experiments. Participants then predicted their partner's experiences on the same measures. The order of the awkwardness items (awkwardness, discomfort) and connectedness items (enjoyment, strength of bond, liking, and perceived understanding) was counterbalanced between pairs.

After reporting their expectations, participants entered the same study room, viewed the discussion questions, and began their conversations. Participants discussed each of the five questions sequentially until they reached their natural conclusions. When finished with their conversation, participants were again separated into individual rooms and reported their experiences in private. Participants first reported their own experiences on the same measures used before the conversation, and then they predicted their partners' responses on the same measures. Participants then completed the ten-item personality inventory (Gosling, Rentfrow, & Swann, 2003).

Finally, participants reported demographic information and were thanked and debriefed.

Results

The awkwardness items ($\alpha_{\text{expectations}} = .91$; $\alpha_{\text{experiences}} = .86$) and connectedness items ($\alpha_{\text{expectations}} = .94$; $\alpha_{\text{experiences}} = .94$) were highly correlated, and so we combined these items to form awkwardness and connectedness scales, respectively.

Manipulation check. To check whether the intimacy manipulation was effective, we recruited a separate group of participants from Amazon's Mechanical Turk ($N = 409$) to rate the intimacy of the control and deep discussion questions. The additional participants confirmed that the manipulation was effective: Participants discussed items that were rated as more intimate in the deep condition ($M = 6.28$, $SD = 1.50$) than in the control condition ($M = 4.74$, $SD = 1.67$), $t(96) = -4.81$, $p < .001$, 95% $CI_{\text{difference}} = [-2.17, -0.90]$, $d = -0.97$ (see Supplemental Material for details).

We then conducted 2 (conversation: control, deep) \times 2 (measurement phase: expectations, experiences) ANOVAs with repeated measures on the second factor and awkwardness, connectedness, and happiness as dependent measures in separate analyses.

Awkwardness. Participants overestimated the awkwardness of their conversations somewhat more in the deep condition than the control condition, as evidenced by a marginally significant conversation \times measurement phase interaction effect, $F(1, 98) = 3.74$, $p = .056$, $\eta_p^2 = .04$ (see Figure 3). Planned contrasts indicated that participants overestimated how awkward their conversation would be both in the control condition, *paired* $t(49) = 4.13$, $p < .001$, 95% $CI_{\text{difference}} = [0.57, 1.65]$, $d = 0.58$, and in the deep condition, *paired* $t(49) = 6.65$, $p < .001$, 95% $CI_{\text{difference}} = [1.30, 2.42]$, $d = 0.94$. However, participants in the deep condition also expected their conversations to feel more awkward than participants in the control condition, $t(98) = -2.26$, $p = .026$, 95% $CI_{\text{difference}} = [-1.82, -0.12]$, $d = -0.45$, even though experiences of awkwardness did not differ significantly between conditions, $t(98) = -0.60$, $p = .552$, 95% $CI_{\text{difference}} = [-0.95, 0.51]$, $d = -0.12$.

Connectedness. Participants underestimated how connected they would feel after speaking with their partner, as indicated by a significant main effect of measurement phase, $F(1,$

98) = 40.46, $p < .001$, $\eta_p^2 = .29$. The conversation \times measurement phase interaction effect was non-significant, $F(1, 98) = 0.01$, $p = .936$, $\eta_p^2 = .0001$. Planned contrasts indicated that although participants underestimated how connected they would feel in both the control, *paired* $t(49) = -4.69$, $p < .001$, 95% $CI_{\text{difference}} = [-1.44, -0.57]$, $d = -0.66$, and deep conditions, *paired* $t(49) = -4.32$, $p < .001$, 95% $CI_{\text{difference}} = [-1.44, -0.52]$, $d = -0.61$, participants in the deep condition did expect to feel more connected to their partner than did participants in the control condition, $t(98) = -3.12$, $p = .002$, 95% $CI_{\text{difference}} = [-1.58, -0.35]$, $d = -0.62$. In this respect, their expectations were calibrated at above-chance levels because participants in the deep condition did indeed feel significantly more connected to their partner than did participants in the control condition, $t(98) = -2.33$, $p = .022$, 95% $CI_{\text{difference}} = [-1.74, -0.14]$, $d = -0.47$.

Happiness. As with the experience of connection, participants also underestimated how happy they would actually feel about their conversations, as indicated by a significant main effect of measurement phase, $F(1, 98) = 45.07$, $p < .001$, $\eta_p^2 = .32$. The conversation \times measurement phase interaction was non-significant, $F(1, 98) = 1.84$, $p = .179$, $\eta_p^2 = .02$. Planned contrasts indicated that participants underestimated how happy they would feel about the conversation in both the control, *paired* $t(49) = -3.79$, $p < .001$, 95% $CI_{\text{difference}} = [-1.33, -0.41]$, $d = -0.54$, and deep conditions, *paired* $t(49) = -5.70$, $p < .001$, 95% $CI_{\text{difference}} = [-1.77, -0.85]$, $d = -0.81$. Participants did not expect to feel differently in the control and deep conditions, $t(98) = -0.73$, $p = .465$, 95% $CI_{\text{difference}} = [-0.85, 0.39]$, $d = -0.15$, but participants in the deep condition reported marginally greater happiness with the conversation when they were finished, $t(98) = -1.72$, $p = .088$, 95% $CI_{\text{difference}} = [-1.44, 0.10]$, $d = -0.34$.



Figure 3. Mean awkwardness, connectedness, and happiness across conversation conditions (control vs. deep) and measurement phase (expectations vs. experiences) in Experiment 3. Error bars $\pm 1 SE$.

Secondary analyses. The magnitude of miscalibration between expectations and experiences did not differ significantly between participants who wrote the final set of discussion questions (Writers) and those who did not (Receivers) on any measures, $F_s(1, 98) \leq 2.28$, $p_s \geq .134$, $\eta_p^2 \leq .02$, suggesting that choosing the topics to discuss does not meaningfully increase the accuracy of people's expectations about the outcomes of the conversation.

As in Experiment 2, our findings were not consistently moderated in any clear way by personality. Overestimating the conversation's awkwardness was not associated with any Big Five traits among participants in the control condition, but was associated with lower openness ($\beta = 0.45$, $t(94) = 4.34$, $p < .001$) among participants in the deep condition. Underestimating connectedness was not associated with any Big Five traits among participants in either the control condition or the deep condition. Underestimating happiness was associated with higher conscientiousness ($\beta = 0.23$, $t(94) = 2.01$, $p = .048$) among participants in the control condition but was not associated with any Big Five traits among participants in the deep condition.

Because we did not observe consistent moderation of any of our results by personality in either Experiments 2 or 3, we did not collect measures of personality in the following experiments.

Discussion

Consistent with Experiments 1 and 2, participants in Experiment 3 overestimated how awkward and uncomfortable a deep conversation with a stranger would be, and also underestimated how connected they would feel to their conversation partner and how happy they would feel about the conversation afterwards. Moreover, we again observed that participants tended to overestimate how awkward a deep conversation would be compared to the conversations they would typically have with a stranger. Although those who were having a deeper-than-normal conversation with a stranger expected the conversation to feel more awkward and uncomfortable than those who were having a typical conversation, the deeper conversations were no more awkward in reality than the typical conversations.

These findings extend our previous experiments in three important ways. First, participants wrote and discussed their *own* questions, thereby replicating the primary results of Experiments 1a-2 across a broad range of ecologically valid conversation topics. Second, the gaps between expectations and experiences in these conversations did not differ meaningfully between participants who chose the discussion topics and those who did not, indicating that people may overestimate the awkwardness of meaningful conversations even when they design the conversations themselves. Third, this experiment suggests that people might have more positive conversations with strangers in daily life if they were willing to dive a little deeper than they normally would.

Finally, these results suggest that miscalibrated expectations, particularly about the awkwardness and discomfort of a deep conversation, may create a barrier to having the deeper

conversations that participants in Experiments 1a-1c reported wishing they had more often. Even though participants in both the control and deep conditions underestimated how connected they would feel to their partner, participants in the deep condition did anticipate feeling more connected than did participants in the control condition. These participants seemed to recognize that a deeper conversation would strengthen their relational bonds with their partner, but mistakenly expected that it would also come with a cost of having a more awkward and uncomfortable conversation. We designed Experiments 4a-7b to test the mechanism underlying these miscalibrated expectations, and to examine how they might encourage conversation that is shallower than participants themselves might consider to be optimal.

Experiments 4a-4b: Underestimating Care

We hypothesized that people undervalue deep conversation at least partly because they underestimate how much others care about, and are interested in, learning the content that will be shared in the conversation itself. People have a fundamental need to feel connected to others (Baumeister & Leary, 1995), a need we are suggesting that people tend to underestimate in others, especially in strangers. Assuming that others are somewhat indifferent to the conversation could lead people to expect that a deep conversation would be especially awkward and uncomfortable, thereby encouraging people to avoid more deep and intimate conversations.

Experiments 1a-1b provided some support for this mechanism, but only included relatively deep conversations. We conducted a more comprehensive test in Experiments 4a and 4b by providing participants with either shallow or deep discussion questions and then asking them to predict how much they would care about the content of their own responses in the interaction, and also how much their partner would care about the participant's own responses. Participants also reported their expectations of how awkward the conversation would feel, how

connected they would feel to their partner, and how happy they would feel about the conversation. After speaking, participants reported their actual experiences on the same measures. We predicted that participants would expect to care about the content of their own responses more than they would expect others to care. We also predicted that participants would underestimate how much others would care about the content of their conversation. Further, we also predicted that underestimating strangers' care and concern would mediate participants' tendency to overestimate awkwardness, but underestimate connectedness and happiness, during the conversation.

Method (Experiment 4a)

Participants. We recruited one hundred and nine pairs of participants during a master's degree student orientation event ($N = 206$ individuals after exclusions; $M_{\text{age}} = 36.83$; $SD_{\text{age}} = 5.48$; 29.61% female; 39.32% Caucasian). We excluded 6 pairs from analyses because 2 pairs completed their conversations before reporting their expectations, 2 pairs included one participant who reported expectations about the wrong set of conversation questions, and 2 pairs included at least one participant who did not consent at the beginning of the session to have their data analyzed in published research. In addition, 14 participants could not be matched with their partner after the session because either 1 or 3 participants reported the same pair ID. We did not analyze data from these participants.

Procedure. The procedure was identical to Experiment 2 with two exceptions. First, we modified the discussion questions and reduced the total number of questions from five to four (due to time constraints in the experimental session). Participants in the shallow condition received the following questions:

1. What do you think about the weather today?

2. What do you think of the city of [removed for blinded review]?
3. How did you celebrate last Halloween?
4. How often do you get your hair cut? Where do you go? Have you ever had a really bad hair cut experience?

Participants in the deep condition received the following questions:

1. For what in your life do you feel most grateful? Tell your partner about it.
2. If a crystal ball could tell you the truth about yourself, your life, your future, or anything else, what would you want to know?
3. If you were going to become a close friend with your partner, please share what would be important for him or her to know.
4. Can you describe a time you cried in front of another person?

Second, we modified the dependent measures. Before meeting their conversation partners, participants reported how much they thought they would care about and feel concerned or interested in their own responses to the questions (0 = *not at all*; 10 = *quite a bit*) and also how much they thought their partner would care about and feel concerned or interested in the participant's responses (0 = *not at all*; 10 = *quite a bit*). These items allow us to test whether participants expect to care more about the content of the conversation than their partner, and if so, whether this self/other caring gap would be especially pronounced in the deep condition. Participants then reported how awkward they would feel during the discussion, how strong of a bond they would feel with their partner, how much they would like their partner, and how happy they would feel about the conversation with their partner on 11-point scales as in the prior

experiments. After reporting their expectations but before speaking, participants indicated how often they wished they would engage in small talk with strangers ($-5 = \text{much less often than I do now}$; $0 = \text{neither less nor more often than I do now}$; $+5 = \text{much more often than I do now}$) and how often they wished they would engage in deep conversations with strangers ($-5 = \text{much less often than I do now}$; $0 = \text{neither less nor more often than I do now}$; $+5 = \text{much more often than I do now}$). Participants then engaged in the conversations and reported their experiences on the same measures used for measuring their expectations.

Finally, participants reported demographic information and were thanked and debriefed.

Results (Experiment 4a)

The strength-of-bond and liking measures were highly correlated in both expectations ($\alpha = .84$) and experiences ($\alpha = .68$) and so we combined these items to form a connectedness scale.

Care measures. As predicted, participants expected to care more about their own responses than their partner would, $F(1, 101) = 32.90, p < .001, \eta_p^2 = .25$. Further, this self/other caring gap was significantly larger in the deep condition than in the shallow condition: A 2 (conversation: shallow, deep) \times 2 (target: self, partner) ANOVA on participants' expectations yielded a significant interaction, $F(1, 101) = 6.00, p = .016, \eta_p^2 = .06$. Although participants in the shallow condition expected to care significantly more about their responses to the questions ($M = 5.69, SD = 1.58$) than their partner would ($M = 5.11, SD = 1.56$), *paired* $t(51) = -2.60, p = .012, 95\% CI_{\text{difference}} = [-1.04, -0.13], d = -0.36$, this gap was significantly larger among participants in the deep condition ($M_s = 7.53$ vs. 6.07 , respectively; $SD_s = 1.24$ vs. 1.41), *paired* $t(50) = -5.26, p < .001, 95\% CI_{\text{difference}} = [-2.02, -0.90], d = -0.74$.

Perhaps more important, comparisons with actual experience indicated that participants systematically underestimated how much they would actually perceive their partner to care about

their responses. A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA on the partner care measure with repeated measures on the second factor produced a significant main effect of measurement phase, $F(1, 101) = 93.22, p < .001, \eta_p^2 = .48$, such that participants underestimated partner care. The conversation \times measurement phase interaction effect was non-significant, $F(1, 101) = 0.29, p = .591, \eta_p^2 = .003$, indicating that participants underestimated the recipients' care similarly in the deep and shallow conditions. Consistent with our hypotheses, participants also underestimated their partner's care significantly more than they underestimated their own, $F(1, 101) = 61.70, p < .001, \eta_p^2 = .38$.

We then conducted a series of 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVAs with repeated measures on the second factor, separately for the awkwardness, connectedness, and happiness measures.

Awkwardness. Participants in the deep condition overestimated how awkward their conversation would feel significantly more than did participants in the shallow condition, as indicated by a significant conversation \times measurement phase interaction, $F(1, 101) = 4.30, p = .041, \eta_p^2 = .04$. Although participants overestimated how awkward the conversation would be in the shallow condition, *paired* $t(51) = 3.76, p < .001, 95\% \text{ CI}_{\text{difference}} = [0.51, 1.70], d = 0.52$, participants overestimated feelings of awkwardness significantly more in the deep condition, *paired* $t(50) = 6.44, p < .001, 95\% \text{ CI}_{\text{difference}} = [1.37, 2.61], d = 0.90$ (see Figure 4, top panel).

Connectedness. Participants underestimated how connected they would feel across conditions, as evidenced by a significant main effect of measurement phase, $F(1, 101) = 144.32, p < .001, \eta_p^2 = .59$. Unexpectedly, the conversation \times measurement phase interaction effect was also significant, $F(1, 101) = 11.76, p < .001, \eta_p^2 = .10$: Although participants underestimated how connected they would feel in the deep condition, *paired* $t(50) = -6.85, p < .001, 95\%$

$CI_{\text{difference}} = [-1.49, -0.81]$, $d = -0.96$, they did so significantly more in the shallow condition, *paired* $t(51) = -9.94$, $p < .001$, $95\% CI_{\text{difference}} = [-2.49, -1.65]$, $d = -1.38$. Participants in the deep condition expected to feel more connected than did participants in the shallow condition, $t(101) = 5.43$, $p < .001$, $95\% CI_{\text{difference}} = [0.89, 1.91]$, $d = 1.07$, and actually did feel more connected in the deep condition than in the shallow condition, $t(101) = 2.46$, $p = .016$, $95\% CI_{\text{difference}} = [0.09, 0.86]$, $d = 0.48$. The significant interaction indicates that these differences in expectation were larger ($d = 1.07$) than the differences in experience ($d = 0.48$).

Happiness. Participants underestimated happiness across conditions, as evidenced by a significant main effect of measurement phase, $F(1, 101) = 130.18$, $p < .001$, $\eta_p^2 = .56$. Unexpectedly, the conversation \times measurement phase interaction effect was also significant for happiness, $F(1, 101) = 13.18$, $p < .001$, $\eta_p^2 = .12$: Although participants significantly underestimated how happy they would feel about the conversation in the deep condition, *paired* $t(50) = -6.09$, $p < .001$, $95\% CI_{\text{difference}} = [-1.36, -0.68]$, $d = -0.85$, they did so significantly more in the shallow condition, *paired* $t(51) = -9.80$, $p < .001$, $95\% CI_{\text{difference}} = [-2.37, -1.57]$, $d = -1.36$. Participants in the deep condition expected to feel happier about their conversation than did participants in the shallow condition, $t(101) = 3.55$, $p < .001$, $95\% CI_{\text{difference}} = [0.42, 1.47]$, $d = 0.70$, but experienced happiness with the conversation did not vary by condition, $t(101) = -0.05$, $p = .962$, $95\% CI_{\text{difference}} = [-0.43, 0.41]$, $d = -0.01$.

Mediational analyses. We found only limited evidence that underestimating a partner's care could explain why participants overestimated how awkward and uncomfortable their interaction would be, while also underestimating how connected and happy they would feel following the conversation. We performed a series of within-pairs mediational analyses using measurement phase (expectation vs. experience) as the independent variable, partner care as the

mediating variable, and awkwardness, connectedness, and happiness as dependent variables in separate analyses. In the shallow condition, the indirect effects were non-significant for awkwardness ($b = 0.19$, 95% CI = [-0.26, 0.64]), non-significant for connectedness ($b = -0.44$, 95% CI = [-1.02, 0.03]), and significant for happiness ($b = -0.43$, 95% CI = [-0.96, -0.03]). In the deep condition, the indirect effects were non-significant for awkwardness ($b = 0.14$, 95% CI = [-0.32, 0.68]), significant for connectedness ($b = -0.47$, 95% CI = [-0.89, -0.17]), and non-significant for happiness ($b = -0.33$, 95% CI = [-0.65, 0.06]). Thus, Experiment 4a provides some evidence that misunderstanding how much others will care explains why deep conversations unfold more positively than people expect, but it does not provide strong evidence of this.

Desire for shallow and deep conversation. Participants reported somewhat stronger desires to engage in deep conversation than shallow conversation with strangers. Participants reported that they wanted to engage in more small talk with strangers than they typically do ($M = 1.12$, $SD = 2.35$), $t(205) = 6.82$, $p < .001$, 95% CI = [0.79, 1.44], $d = 0.47$, and also that they wanted to engage in more deep conversation with strangers than they typically do ($M = 1.48$, $SD = 2.12$), $t(205) = 9.99$, $p < .001$, 95% CI = [1.18, 1.77], $d = 0.70$. Participants' preference to engage in more deep conversation was marginally stronger than their preference to engage in more small talk, *paired* $t(205) = -1.90$, $p = .059$, 95% CI_{difference} = [-0.73, 0.01], $d = -0.13$. These results are important because they indicate that our participants would generally prefer interacting with others more, rather than less, suggesting that any reluctance to engage in deeper conversation than one normally would does not stem from participants' own disinterest in doing so. These preference ratings instead suggest a psychological barrier that may keep people from engaging as deeply with others as much as they might prefer.

Method & Results (Experiment 4b)

We conducted a replication experiment with only minor changes to the procedure ($N = 101$ pairs; see Supplemental Material for the full method and results). Experiment 4b largely replicated the results of Experiment 4a: Participants overestimated the awkwardness of their conversations marginally more in the deep condition than the shallow condition, $F(1, 99) = 3.69$, $p = .058$, $\eta_p^2 = .04$, but underestimated their connectedness significantly more in the shallow condition than the deep condition, $F(1, 99) = 17.96$, $p < .001$, $\eta_p^2 = .15$, and underestimated their happiness marginally more in the shallow condition, $F(1, 99) = 3.21$, $p = .076$, $\eta_p^2 = .03$ (see Figure 4, bottom panel).

Experiment 4b also found somewhat stronger evidence that underestimating a partner's care could explain why people undervalue the positive outcomes of their conversations. We conducted mediational analyses with measurement phase (expectation vs. experience) as the independent variable, partner care as the mediating variable, and each of the primary measures as dependent variables in separate analyses. In the shallow condition, the indirect effects were non-significant for awkwardness ($b = 0.35$, 95% CI = [-0.31, 0.99]), but significant for connectedness ($b = -0.54$, 95% CI = [-1.02, -0.19]), and happiness ($b = -0.36$, 95% CI = [-0.63, -0.17]). In the deep condition, the indirect effects were likewise non-significant for awkwardness, ($b = 0.37$, 95% CI = [-0.25, 0.92]), but significant for connectedness ($b = -0.53$, 95% CI = [-0.93, -0.24]), and happiness ($b = -0.39$, 95% CI = [-0.89, -0.06]).

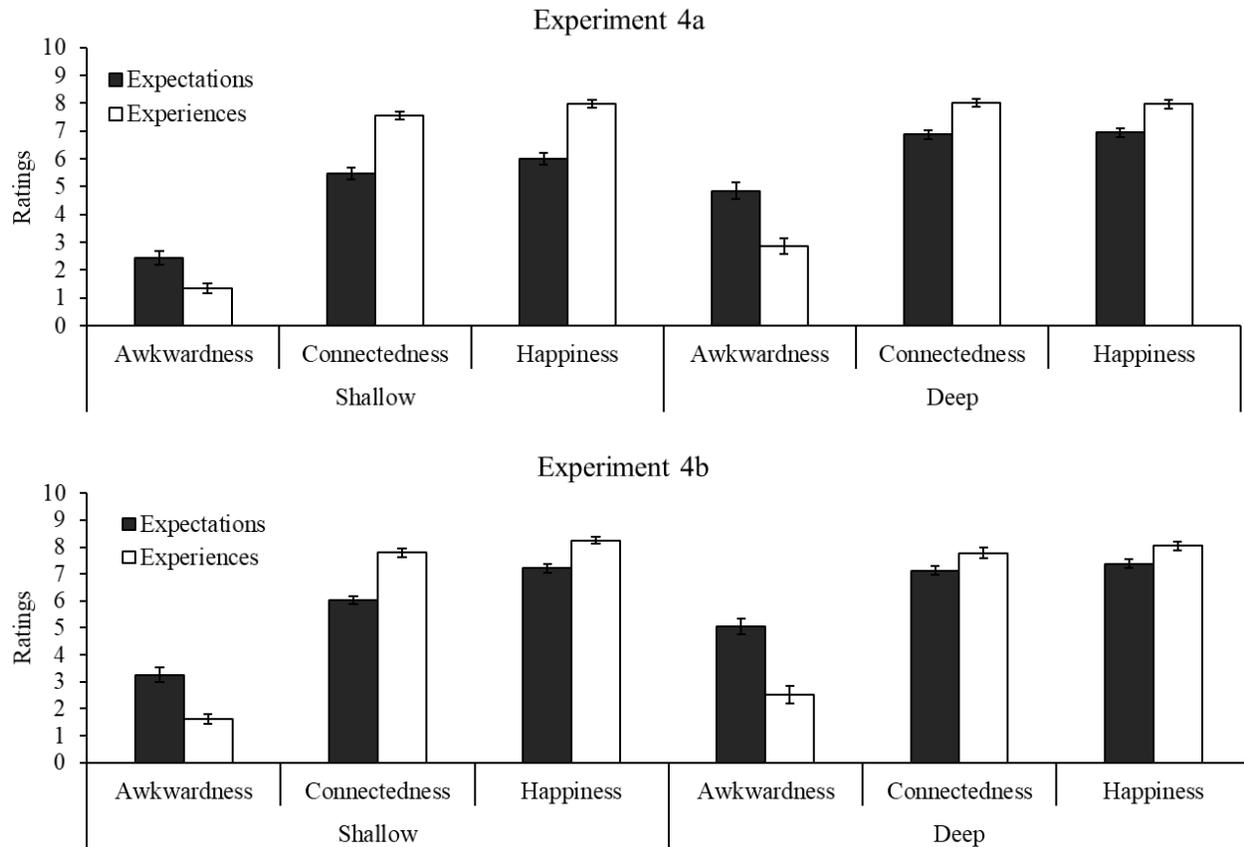


Figure 4. Mean awkwardness, connectedness, and happiness across conversation conditions (shallow vs. deep) and measurement phase (expectations vs. experiences) in Experiments 4a and 4b. Error bars ± 1 SE.

Discussion (Experiments 4a-4b)

Participants underestimated how positive they would feel in conversation with a stranger, expecting to feel more awkward and less connected and happy than they actually did. Consistent with the prior experiments, participants overestimated how awkward the conversation would feel to a greater extent in the deep conditions than in the shallow conditions. Importantly, participants underestimated how much others would care about the content of the conversation, but we found

only mixed evidence that underestimating strangers' care helped to explain why participants underestimated the overall positivity of their conversations.

These experiments also revealed an unexpected pattern of results that we did not observe in Experiments 2 and 3. Although participants in the shallow condition expected to feel substantially less connected to one another than participants in the deep condition, participants in both conditions felt strongly connected after speaking. As a result, participants in the shallow condition underestimated connectedness *more* than participants in the deep condition. Anecdotal reports during the session debriefing indicated that participants in the shallow condition often turned what we intended to be relatively superficial conversations into deeper and more meaningful conversations than they had anticipated, meaning that our experimental manipulations were not as strong as intended. Although evaluations of intimacy differed significantly between the deep ($M = 7.18$, $SD = 1.61$) and shallow conditions ($M = 5.92$, $SD = 1.84$) in Experiment 4b, $t(64) = 2.95$, $p = .004$, $95\% CI_{\text{difference}} = [0.41, 2.11]$, $d = 0.73$, intimacy ratings even in the shallow condition exceeded the midpoint of the scale, *one-sample* $t(32) = 2.88$, $p = .007$, $95\% CI = [5.27, 6.58]$, $d = 0.50$. Instead of being limited to the questions that we provided, which would have guided people's expectations, actual conversations seemed to drift in a more meaningful direction than people may have anticipated based on the questions alone. We think this is a potentially interesting and unexpected consequence in some conversations, which might help to explain why conversations with strangers can be surprisingly positive (Epley & Schroeder, 2014). We will discuss this possibility in greater detail in the General Discussion.

Experiment 5: Manipulating Perceived Care Via Relationships

Experiments 1a-4b suggest that people expect deeper conversations to create stronger connections than shallow conversations, but also that deeper conversations will be relatively awkward and uncomfortable. These experiments also indicate that these expectations are somewhat miscalibrated, such that people underestimate how connected they will feel to their partners and overestimate how awkward and uncomfortable their deep conversations will be. However, Experiments 4a and 4b provide only mixed evidence that miscalibrated expectations about the consequences of deep conversation stem from underestimating others' care and interest in learning more intimate information about oneself. Perhaps more important, correlational analyses of mediation do not test the causal influence of a mediator, which only an experimental approach to testing mediation can do. We therefore designed Experiment 5 to provide a causal test of the importance of a partner's perceived care and interest by manipulating it directly through the strength of an existing social relationship. Specifically, we asked participants to have a conversation with either a distant stranger or with a close other who would already be known to care about one's responses to intimate conversation topics, including a friend, family member, or romantic partner. We predicted that participants would be less likely to underestimate a close other's interest compared to a distant stranger's interest, and would also be more calibrated when predicting feelings of awkwardness, enjoyment, and happiness before speaking with a close other compared to a distant stranger.

We also designed Experiment 5 to address a potential methodological concern. In the previous experiments, participants did not see their partner before reporting their expectations, and so it is possible that they imagined meeting an atypical stranger who would be even less interested in the conversation than the stranger they were later paired with. We theorize instead that people underestimate how much even a *typical* stranger will care about the content of a

conversation. To address this concern, participants in Experiment 5 were able to see their partner in the stranger condition at the start of the study session. We still predicted, consistent with the previous experiments, that participants in the “stranger” condition would underestimate their partner’s interest and therefore underestimate how positively their conversations would unfold.

Method

Participants. We targeted 200 pairs of participants and finished recruiting once that target was reached after data exclusions. We achieved this by recruiting 204 pairs of participants from several public parks ($N = 400$ individuals after exclusions; $M_{\text{age}} = 35.27$; $SD_{\text{age}} = 16.21$; 61.50% female; 61.50% Caucasian) to complete the study in exchange for a \$5 gift card. We excluded 4 of these pairs from analyses because both participants in one pair answered their phones during the conversation, and because participants in three pairs discussed the dependent measures while responding to those measures. Among 200 participants in the “close other” condition, 31.50% reported that they were friends, 24.50% reported that they were spouses, 21.50% reported that they were dating, 17.00% reported that they were family members, 3.00% reported that they were acquaintances, 2.00% reported that they were colleagues, and 0.50% did not report the nature of their relationship.

Procedure. Experimenters recruited either pairs of distant strangers who had never met one another or pairs of close friends, family, or partners who were visiting the park together. Participants in both the close and distant conditions saw their conversation partner at the beginning of the experiment before receiving instructions. After providing informed consent, participants in the close conditions reported how close or connected they currently felt to the other person (0 = *not close at all*; 10 = *extremely close*).

In both conditions, participants were separated and were then told by different experimenters that they would develop a set of discussion questions that they might later ask and answer with the other participant. Following a similar procedure to that used in Experiment 3, participants first generated two control questions, and then were asked to generate two deeper questions. The instructions in the distant stranger were nearly identical to those used in Experiment 3, while the instructions in the close other condition asked participants to write questions that they typically discuss with this person (control) or that were deeper than they normally discuss with this person (see the Supplemental Material for complete instructions).

The procedure was then largely identical to that of Experiment 3, except that we included slightly different dependent variables given that participants in the close condition were already acquainted and connected with each other. In particular, participants first reported how much they would care about and feel concerned or interested in their own responses, and how much their partner would care about and feel concerned or interested in the participant's responses, on separate scales. Participants then reported how awkward they would feel during the discussion, how much they would enjoy the conversation, and how happy they would feel about the conversation, on 11-point scales as in the prior experiments. The awkwardness and enjoyment items were counterbalanced between pairs. As a manipulation check, participants then viewed the control questions and the deep questions that they had written earlier—regardless of whether their own questions or their partner's questions were selected for the discussion—and separately rated the intimacy of each pair of questions (0 = *not intimate at all*; 10 = *extremely intimate*).

Participants then viewed the discussion questions and began their conversations as in Experiment 3, and then reported their experiences on the same measures used to measure their

expectations before the conversation. As a second manipulation check, participants also reported how intimate their conversation was (0 = *not at all intimate*; 10 = *extremely intimate*).

Finally, participants reported demographic information and received their compensation.

Results

Manipulation checks. The intimacy manipulation was effective. Participants reported that their questions were less intimate in the control conditions than in the deep conditions, $F(1, 196) = 214.30, p < .001, \eta_p^2 = .52$, and this pattern did not differ depending on whether they were distant strangers or close others, $F(1, 196) = 0.61, p = .436, \eta_p^2 = .003$. After the conversation, participants likewise reported having less intimate conversations in the control conditions than in the deep conditions, $F(1, 196) = 63.05, p < .001, \eta_p^2 = .24$, again regardless of whether they were distant strangers or close others, $F(1, 196) = 0.20, p = .658, \eta_p^2 = .001$. Participants also reported having deeper conversations with close others than with distant strangers, $F(1, 196) = 10.48, p = .001, \eta_p^2 = .05$.

Care measures. As anticipated, participants expected to care about their own responses more than their partner would across experimental conditions, $F(1, 196) = 6.73, p = .010, \eta_p^2 = .03$. Furthermore, this self/other caring gap was significantly larger in the deep conditions than in the control conditions, as we observed in Experiment 4. Specifically, a 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (target: self, partner) ANOVA on participants' expectations yielded a significant conversation \times target interaction, $F(1, 196) = 8.86, p = .003, \eta_p^2 = .04$. Participants in the deep conditions expected to care more about their responses to the questions than their partner would, $F(1, 98) = 19.11, p < .001, \eta_p^2 = .16$, but participants in the control conditions did not, $F(1, 98) = 0.06, p = .805, \eta_p^2 = .001$.

More important for the unique contribution of this experiment, pairs in the distant conditions underestimated each other's care more than did pairs in the close conditions. A 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVA on the partner care measure with repeated measures on the third factor produced a significant relationship \times measurement phase interaction effect, $F(1, 196) = 55.84, p < .001, \eta_p^2 = .22$. Although participants who spoke with close others underestimated how much their partner would care about their responses, $F(1, 98) = 4.13, p = .045, \eta_p^2 = .04$, those who spoke with distant strangers did so significantly more, $F(1, 98) = 156.61, p < .001, \eta_p^2 = .62$.

We next report a series of 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVAs with repeated measures on the third factor, separately for the awkwardness, enjoyment, and happiness measures.

Awkwardness. Participants in the deep conditions overestimated how awkward their conversations would be more than did participants in the control conditions, as revealed by a significant conversation \times measurement phase interaction effect, $F(1, 196) = 4.94, p = .027, \eta_p^2 = .02$. Importantly, we also observed a significant relationship \times measurement phase interaction effect, $F(1, 196) = 11.05, p = .001, \eta_p^2 = .05$ (see Figure 5): Although participants in the close conditions overestimated how awkward and uncomfortable their conversations would feel, $F(1, 98) = 19.36, p < .001, \eta_p^2 = .16$, participants in the distant conditions did so significantly more, $F(1, 98) = 86.34, p < .001, \eta_p^2 = .47$.

Analyses of simple effects within this interaction indicated that among distant strangers, participants in the deep condition expected that their conversations would feel marginally more awkward than those in the control condition, $t(98) = -1.88, p = .062, 95\% \text{ CI}_{\text{difference}} = [-1.46,$

0.04], $d = -0.38$, but did not feel significantly different in experience ($M_s = 1.49$ vs. 1.56, respectively; $SD_s = 1.50$ vs. 1.52), $t(98) = 0.23$, $p = .817$, 95% $CI_{\text{difference}} = [-0.53, 0.67]$, $d = 0.05$. Among close others, participants in the deep condition likewise expected their conversations to feel more awkward than did those in the control condition, $t(98) = -2.58$, $p = .011$, 95% $CI_{\text{difference}} = [-1.68, -0.22]$, $d = -0.52$, but felt only marginally more awkward in actual experience ($M_s = 1.88$ vs. 1.20, respectively; $SD_s = 2.10$ vs. 1.52), $t(98) = -1.86$, $p = .066$, 95% $CI_{\text{difference}} = [-1.41, 0.05]$, $d = -0.37$. Once again, deep conversations were not as awkward and uncomfortable as expected.

Enjoyment. As predicted, participants underestimated how much they would enjoy their conversations across conditions, as indicated by a significant main effect of measurement phase, $F(1, 196) = 135.31$, $p < .001$, $\eta_p^2 = .41$. More important, we also observed a significant relationship \times measurement phase interaction effect, $F(1, 196) = 29.58$, $p < .001$, $\eta_p^2 = .13$: Although participants in the close conditions underestimated how much they would enjoy their conversations, $F(1, 98) = 21.45$, $p < .001$, $\eta_p^2 = .18$, participants in the distant conditions underestimated their enjoyment significantly more, $F(1, 98) = 131.76$, $p < .001$, $\eta_p^2 = .57$. Neither the main effect of conversation, nor interactions with conversation, were significant, $F_s(1, 196) \leq 2.15$, $p_s \geq .144$, $\eta_p^2_s \leq .01$, indicating that participants expected and experienced similar enjoyment in the control and deep conditions.

Happiness. Participants underestimated how happy they would feel about their conversations across conditions, as indicated by a significant main effect of measurement phase, $F(1, 196) = 188.47$, $p < .001$, $\eta_p^2 = .49$. More important, we also observed a significant relationship \times measurement phase interaction effect, $F(1, 196) = 29.96$, $p < .001$, $\eta_p^2 = .13$: Although participants in the close conditions underestimated how happy they would feel about

their conversations, $F(1, 98) = 34.31, p < .001, \eta_p^2 = .26$, participants in the distant conditions did so significantly more, $F(1, 98) = 183.11, p < .001, \eta_p^2 = .65$. Neither the main effect of conversation, nor interactions with conversation, were significant, $F_s(1, 196) \leq 3.28, p_s \geq .072, \eta_p^2_s \leq .02$, indicating that participants expected and experienced similar levels of happiness in the control and deep conditions.

Mediational analyses. We predicted that participants would overestimate awkwardness, and underestimate their enjoyment and happiness, because they would underestimate how much their partner would care about one's responses. We did not observe support for these hypotheses in the control conditions, but observed stronger support in the deep conditions.

Specifically, we performed a series of within-pairs mediational analyses with measurement phase (expectation vs. experience) as the independent variable and partner care as the mediating variable. Among participants in the control condition who spoke with distant strangers, the indirect effects were non-significant for awkwardness ($b = 0.25, 95\% \text{ CI} = [-0.25, 0.82]$), non-significant for enjoyment ($b = -0.30, 95\% \text{ CI} = [-0.76, 0.19]$), and non-significant for happiness ($b = -0.31, 95\% \text{ CI} = [-0.63, 0.05]$). Among participants in the control condition who spoke with close others, the indirect effects were non-significant for awkwardness ($b = 0.02, 95\% \text{ CI} = [-0.03, 0.24]$), non-significant for enjoyment ($b = -0.07, 95\% \text{ CI} = [-0.41, 0.09]$), and non-significant for happiness ($b = -0.06, 95\% \text{ CI} = [-0.29, 0.11]$).

Among participants in the deep condition who spoke with distant strangers, the indirect effects were significant for awkwardness ($b = 0.88, 95\% \text{ CI} = [0.36, 1.44]$), non-significant for enjoyment ($b = -0.28, 95\% \text{ CI} = [-1.13, 0.51]$), and significant for happiness ($b = -0.46, 95\% \text{ CI} = [-0.98, -0.12]$). Among participants in the deep condition who spoke with close others, the indirect effects were non-significant for awkwardness ($b = -0.004, 95\% \text{ CI} = [-0.25, 0.19]$),

significant for enjoyment ($b = -0.21$, 95% CI = [-0.50, -0.03]), and significant for happiness ($b = -0.14$, 95% CI = [-0.36, -0.02]). People may underestimate the positivity of deep conversations in part because their conversation partners are more caring and interested than people anticipate.

Furthermore, underestimating others' care helped to explain why participants who spoke with distant strangers were more likely to underestimate enjoyment and happiness than those who spoke with close others. We performed a series of between-pairs mediational analyses with relationship (close vs. distant) as the independent variable, underestimation of partner care as the mediating variable, and the magnitude of miscalibration for each of the primary measures as dependent variables in separate analyses, using the PROCESS macro (Hayes, 2013). Among participants in the control conditions, the indirect effects were non-significant for awkwardness ($b = 0.27$, 95% CI = [-0.04, 0.69]), significant for enjoyment ($b = -0.57$, 95% CI = [-0.97, -0.16]), and significant for happiness ($b = -0.54$, 95% CI = [-0.88, -0.24]). Among participants in the deep conditions, the indirect effects were also non-significant for awkwardness ($b = 0.34$, 95% CI = [-0.004, 0.78]), significant for enjoyment ($b = -0.41$, 95% CI = [-0.91, -0.003]), and significant for happiness ($b = -0.37$, 95% CI = [-0.71, -0.14]). People's expectations of conversations with close others may be more calibrated because they more correctly recognize how much close others will care about the content of their conversation.

Secondary analyses. Consistent with Experiment 3, the primary results did not differ meaningfully between Writers (who wrote the final set of discussion questions) and Receivers (who did not). A series of 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) \times 2 (role: Writer, Receiver) ANOVAs on awkwardness, enjoyment, and happiness with repeated measures on the third and fourth factors yielded non-significant measurement phase \times role interaction effects for awkwardness, $F(1, 196)$

= 3.50, $p = .063$, $\eta_p^2 = .02$, enjoyment, $F(1, 196) = 1.94$, $p = .166$, $\eta_p^2 = .01$, and happiness, $F(1, 196) = 0.31$, $p = .578$, $\eta_p^2 = .002$. People seem to underestimate the positive outcomes of their conversations even when they generate the topics themselves.

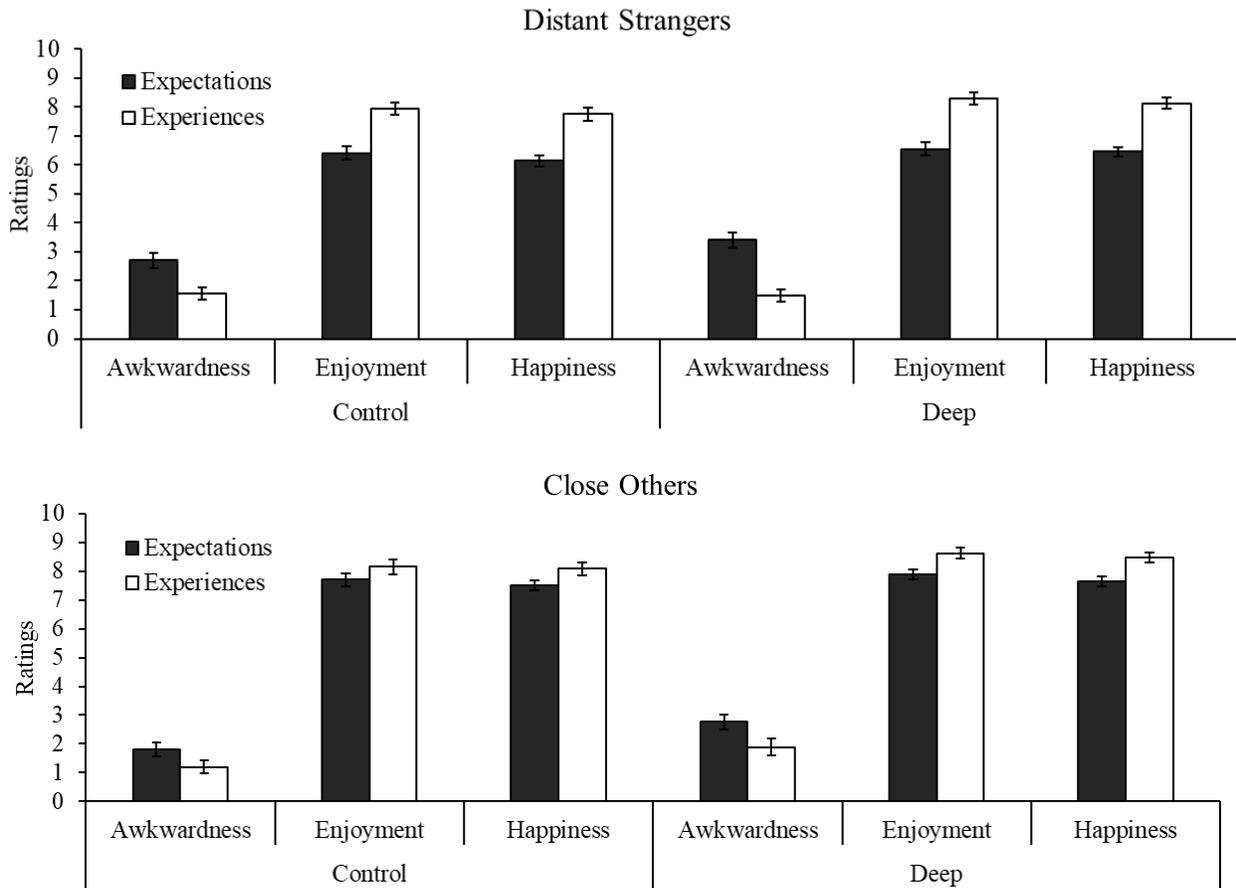


Figure 5. Mean awkwardness, enjoyment, and happiness across conversation conditions (control vs. deep), relationship (distant vs. close) and measurement phase (expectations vs. experiences) in Experiment 5. Error bars $\pm 1 SE$.

Discussion

Experiment 5 reveals that people are more calibrated anticipating how much close friends, family members, and partners will care about their conversation. As a result, people are also more calibrated predicting how awkward, enjoyable, and happy they will feel in the

conversation with close others compared to with distant strangers. As we observed in prior experiments, people overestimate how awkward deep conversations with strangers will be more than typical conversations, but underestimate how enjoyable and happy they will feel about both typical and deeper conversations. We did not measure connectedness in Experiment 5, which sometimes varied by conversation content in prior experiments, because we directly manipulated social connection across conditions in this experiment.

Because people expect close others to be relatively more caring and interested than distant strangers, our theory also predicted that people would prefer to have deeper conversations with close others. The data from Experiment 5 allow us to begin testing this hypothesis because participants in the control conditions discussed questions that they would typically discuss with their conversation partner (close others) or that they typically discuss while getting to know somebody new (distant strangers). In these control conditions, participants who spoke with close others ($M = 3.70$, $SD = 2.65$) reported that their conversation topics were more intimate than those who spoke with distant strangers ($M = 2.50$, $SD = 2.40$), $t(98) = 2.37$, $p = .020$, 95% $CI_{\text{difference}} = [0.20, 2.20]$, $d = 0.47$. After the conversation, participants who spoke with close others ($M = 5.03$, $SD = 2.46$) also reported marginally more intimate conversations than those who spoke with distant strangers ($M = 4.21$, $SD = 2.22$), $t(98) = 1.75$, $p = .083$, 95% $CI_{\text{difference}} = [-0.11, 1.75]$, $d = 0.35$. People may engage in deeper conversations with others who they perceive to be caring and interested. Underestimating distant strangers' care may therefore create a barrier to having deeper and more intimate conversations.

Of course, Experiment 5 was not designed to test whether people prefer to have deeper conversations with close others than distant strangers. We therefore conducted a supplementary experiment on Amazon's Mechanical Turk ($N = 109$) in which participants imagined visiting the

lab for a conversation with either a close friend or a distant stranger. Participants viewed a list of 20 pre-tested discussion questions (see Appendix) and selected the 5 questions they preferred to discuss with the other person. Participants in the “friend” condition ($M = 3.10$, $SD = 1.51$) chose a greater number of deep questions than participants in the “stranger” condition ($M = 1.79$, $SD = 1.32$), $t(107) = 4.82$, $p < .001$, 95% CI = [0.77, 1.84], $d = 0.92$, and likewise selected questions higher in average intimacy ($M = 5.04$, $SD = 1.24$) than did participants in the “stranger” condition, ($M = 3.97$, $SD = 1.10$), $t(107) = 4.76$, $p < .001$, 95% CI_{difference} = [0.62, 1.51], $d = 0.91$. Furthermore, participants who imagined speaking with a close friend reported significantly greater interest in discussing the deep questions ($M = 7.00$, $SD = 1.64$), $t(107) = 4.28$, $p < .001$, 95% CI_{difference} = [0.83, 2.26], $d = 0.82$, and expected the other person to care more about one’s own responses to the deep questions ($M = 7.11$, $SD = 1.42$), $t(107) = 3.67$, $p < .001$, 95% CI_{difference} = [0.56, 1.88], $d = 0.70$, compared to participants who imagined speaking with a distant stranger (M s = 5.45, 5.89, respectively; SD s = 2.08, 1.98; see Supplemental Material for the full method and results).

These findings provide further evidence that people refrain from having deep and intimate conversations when they are concerned that another person will be uncaring and indifferent toward the conversation, and thus suggest that underestimating strangers’ care may create a psychological barrier to having deeper conversations with distant others. Experiments 7a-7b will test whether removing this barrier encourages people to choose deeper conversation topics.

Experiments 6a-6b: Engaging in Shallow *and* Deep Conversations

Our data thus far reveal that participants overestimate the awkwardness of deep conversations more than shallow ones, and expect a wide range of conversations to lead to

weaker connections and less happiness than they do. However, two limitations of the previous designs have hindered our ability to test for differences in experiences between shallow and deep conversations. We address both limitations in the following experiments. First, participants have rated their experiences near the ceilings of our connectedness scales: Participants tend to feel highly connected to their conversation partner after a shallow *or* deep conversation. We address this limitation in Experiments 6a and 6b by instructing participants to engage in a shallow conversation with one stranger and a deep conversation with another stranger, enabling clearer comparisons of their connections to two different partners. We hypothesized that participants would report feeling more connected to their deep conversation partner.

Second, in the prior experiments, we included both positive measures of partner care, connectedness, and happiness, and negative measures of awkwardness. People perceive their conversation partner to be more caring during deep conversations, and at times feel happier and more connected to the person, but they have also felt more awkward during deep conversations in some experiments. This mix of positive and negative experiences does not indicate whether participants themselves prefer their overall experiences in deep conversation relative to shallow conversation, nor whether engaging in these conversations would affect their preference for future shallow versus deep conversations. To address these limitations, participants in Experiment 6b engaged in both types of conversations and then indicated which conversation they preferred. We hypothesized that participants would tend to prefer their deep conversations.

Method (Experiment 6a)

Participants. We recruited 30 groups of four participants ($N = 120$ individuals; $M_{\text{age}} = 28.33$; $SD_{\text{age}} = 11.89$; 65.00% female; 39.17% Caucasian). Among these, 10 groups were recruited during an event advertised as “Dinner and a Study Night” in exchange for \$10 and

dinner, and 20 groups were recruited through the university's Virtual Lab using the Zoom video conferencing software, in exchange for an \$8 digital gift card, due to the COVID-19 pandemic. Two groups were excluded from analyses: In one group, the participants engaged in their second conversation before reporting expectations, and in the other, Internet connection problems prevented the participants from finishing their conversations (see Supplemental Material for the analyses including these groups). In addition, another six groups were removed from analyses and were not analyzed: In 4 groups, one participant typed in the wrong condition code and reported expectations about the wrong topics before each conversation, and in 2 groups, one participant typed in the wrong group ID and could not be matched in the data with their group members.

Procedure. Participants were randomly assigned to one of four roles (A, B, C, D) to determine who they would speak with during each round of conversation. Participants were then informed that they were about to speak with another person who they had never met before. They then read the three conversation questions. In the shallow-first counterbalance, participants read the following questions:

1. How is your day going so far?
2. What do you think about the weather today?
3. How often do you come here?⁴

In the deep-first counterbalance, participants instead read the following questions:

1. For what in your life do you feel most grateful? Please tell me about it.

⁴ After moving the study to our "Virtual Lab" during the COVID-19 pandemic, we changed the third conversation question to, "How often do you participate in research studies?"

2. If you could undo one mistake you have made in your life, what would it be and why would you undo it?
3. Can you tell me about one of the last times you cried in front of another person?

Participants then reported their expectations about the conversation, including how much they would care about and feel concerned or interested in their own responses, how much their partner would care about and feel concerned or interested in one's responses, how awkward they would feel while speaking with the other person, how strong of a bond they would experience with the other person, how much they would like the other person, and how happy they would feel after the discussion, on 11-point scales as in the prior experiments.

Participants were then paired with another group member (A with D; B with C) for the discussion. Because anecdotal evidence suggested that participants in the prior experiments discussed content outside the scope of the conversation questions, potentially weakening the intimacy manipulations, the experimenter instructed participants not to drift off-topic while discussing the questions. After their conversations, participants reported their actual experiences on the same measures described above. Participants also reported how intimate the conversation was (0 = *light conversation and small talk*; 10 = *meaningful conversation and deep talk*) and how negative or positive the topics they discussed were (-5 = *very negative*; 0 = *equally positive and negative*; 5 = *very positive*).

Participants were then informed that they would discuss a new set of conversation questions with a different person who they hadn't met before. Participants in the shallow-first counterbalance now read the deep questions; those in the deep-first counterbalance now read the shallow questions. Participants reported their expectations, engaged in their conversations (A

with C; B with D), and reported their experiences on the same measures described above. After finishing the second conversation, participants additionally reported which person they felt more connected to ($-5 = I$ feel much more connected to the *FIRST* person; $0 = I$ feel equally connected to *BOTH* people; $5 = I$ feel much more connected to the *SECOND* person), then indicated how intimate their typical conversations with strangers are versus how intimate they wished these conversations would be on separate scales. Participants recruited to the Virtual Lab then indicated whether they had trouble seeing or hearing the other participants during their conversations over Zoom.

Finally, participants completed demographic items and were thanked and debriefed.

Method (Experiment 6b)

Participants. We recruited 30 groups of four participants ($N = 120$ individuals; $M_{\text{age}} = 26.01$; $SD_{\text{age}} = 9.15$; 61.67% female; 25.83% Caucasian) in exchange for a \$7 digital gift card through the university's Virtual Lab.

Procedure. The procedure was identical to Experiment 6a with several exceptions. First, immediately after providing informed consent, participants immediately read both sets of conversation questions and then indicated which conversation they thought they would prefer ($-5 = \textit{will strongly prefer Conversation A}$; $0 = \textit{will equally prefer both conversations}$; $5 = \textit{will strongly prefer Conversation B}$). Second, we removed the remaining expectation items before each conversation (care, awkwardness, connectedness, happiness). We did this because participants in our prior experiments may have anchored their experienced awkwardness ratings on their earlier expectations, potentially exaggerating differences in actual awkwardness between shallow and deep conversations. Third, we included three additional experience items after each conversation in which participants indicated their interest in being friends with the person they

spoke with: specifically, how interested they would be in speaking with the other person again outside the lab (0 = *not at all*; 10 = *very*), how much they would have to talk about with this person outside the lab (0 = *nothing at all*; 10 = *quite a bit*), and how strong a friend they would become upon getting to know this person outside the lab (0 = *weak, like a stranger*; 10 = *strong, like a close friend*).

Finally, after reporting their experiences for both conversations, participants completed two items about their preferences for shallow versus deep conversation. They first thought about the two conversations they just engaged in and indicated which conversation they preferred (-5 = *strongly preferred the FIRST conversation*; 0 = *equally preferred both conversations*; 5 = *strongly preferred the SECOND conversation*). They then imagined participating in a future experiment in which they would discuss the same two sets of questions with two different strangers who they had never met before. Participants reread both sets of questions and indicated which future conversation they thought they would prefer (-5 = *will strongly prefer Conversation A*; 0 = *will equally prefer both conversations*; 5 = *will strongly prefer Conversation B*).

Results (Experiment 6a)

The strength-of-bond and liking items were highly correlated for both the shallow ($\alpha_{\text{expectations}} = .77$, $\alpha_{\text{experiences}} = .72$) and deep ($\alpha_{\text{expectations}} = .80$, $\alpha_{\text{experiences}} = .68$) conversations, and so we collapsed these items to form a connectedness scale. The results described below were not qualified by significant interactions with study location (dinner event vs. Virtual Lab) and so we pooled the data across both locations.

Manipulation check. The intimacy manipulation was effective: Participants reported that their deep conversations ($M = 6.37$, $SD = 2.19$) were significantly more intimate than their

shallow conversations ($M = 3.08$, $SD = 2.64$), $paired\ t(119) = -11.84$, $p < .001$, $95\% CI_{\text{difference}} = [-3.84, -2.74]$, $d = -1.35$.

Care measures. Replicating the prior experiments, participants expected to care more about their own responses than their partner would, $b = 1.51$, $SE = 0.14$, $t(298.19) = 10.49$, $p < .001$, $95\% CI = [1.23, 1.79]$. Further, this self/other caring gap was significantly larger for deep conversations than shallow conversations. A mixed linear model with fixed-effects terms for conversation (shallow vs. deep), target (own care vs. partner care), and their interaction, and random intercepts for the participant and the partner, yielded a significant intimacy \times target interaction, $b = 0.97$, $SE = 0.29$, $t(298.19) = 3.36$, $p < .001$, $95\% CI = [0.40, 1.53]$: Although participants expected to care significantly more about their responses during shallow conversations ($M = 5.05$, $SD = 2.61$) than their partner would ($M = 4.03$, $SD = 2.39$), $b = 1.03$, $SE = 0.20$, $t(298.19) = 5.04$, $p < .001$, $95\% CI = [0.63, 1.42]$, this gap was significantly larger for deep conversations ($M_s = 7.23$ vs. 5.24 , respectively; $SD_s = 2.04$ vs. 2.21), $b = 1.99$, $SE = 0.20$, $t(298.19) = 9.79$, $p < .001$, $95\% CI = [1.59, 2.39]$.

Perhaps more important, comparisons with actual experience indicated that participants systematically underestimated how much they would actually perceive their partner to care about their responses. A mixed linear model with fixed-effects terms for conversation (shallow vs. deep), measurement phase (expectations vs. experiences), and their interaction, and random intercepts for the participant and the partner, yielded a significant effect of measurement phase, $b = 1.77$, $SE = 0.16$, $t(357.00) = 11.25$, $p < .001$, $95\% CI = [1.46, 2.07]$, such that participants underestimated partner care. The conversation \times measurement phase interaction effect was non-significant, $b = -0.38$, $SE = 0.31$, $t(357.00) = -1.22$, $p = .223$, $95\% CI = [-1.00, 0.23]$, indicating that participants underestimated the recipients' care similarly for the deep and shallow

conversations. Consistent with our hypotheses, participants also underestimated their partners' care significantly more than they underestimated their own, $b = 1.42$, $SE = 0.21$, $t(773.31) = 6.61$, $p < .001$, 95% CI = [1.00, 1.84].

We then constructed a series of mixed linear models with fixed-effects terms for conversation, measurement phase, and their interaction, and random intercepts for the participant and the partner, separately for the awkwardness, connectedness, and happiness measures.

Awkwardness. Participants again overestimated the awkwardness of deep conversations significantly more than shallow ones, as indicated by a significant conversation \times measurement phase interaction, $b = 0.94$, $SE = 0.38$, $t(302.11) = 2.48$, $p = .014$, 95% CI = [0.20, 1.69].

Although participants overestimated how awkward their shallow conversations would be, $b = -0.93$, $SE = 0.27$, $t(302.11) = -3.44$, $p < .001$, 95% CI = [-1.45, -0.40], they did so significantly more for deep conversations, $b = -1.87$, $SE = 0.27$, $t(302.11) = -6.94$, $p < .001$, 95% CI = [-2.39, -1.34] (see Figure 6).

Connectedness. As predicted, participants underestimated how connected they would feel to both partners, as shown by a significant effect of measurement phase, $b = 1.02$, $SE = 0.18$, $t(292.99) = 5.58$, $p < .001$, 95% CI = [0.66, 1.38]. The conversation \times measurement phase interaction was non-significant, $b = -0.16$, $SE = 0.26$, $t(292.99) = -0.61$, $p = .541$, 95% CI = [-0.66, 0.35]: Participants expected to feel more connected to their deep conversation partner, $b = 0.98$, $SE = 0.18$, $t(292.99) = 5.38$, $p < .001$, 95% CI = [0.63, 1.34], and felt significantly more connected to this person after speaking, $b = 1.14$, $SE = 0.18$, $t(292.99) = 6.24$, $p < .001$, 95% CI = [0.78, 1.50].

We also analyzed the item in which participants directly compared the strength of their connection to their two conversation partners. Participants reported significantly stronger

connections to their deep conversation partner than their shallow conversation partner ($M = 1.62$, $SD = 3.00$), *one-sample* $t(119) = 5.90$, $p < .001$, 95% CI = [1.07, 2.16], $d = 0.54$.

Happiness. As predicted, participants underestimated how happy they would feel after both conversations, as indicated by a significant effect of measurement phase, $b = 1.17$, $SE = 0.17$, $t(296.00) = 7.03$, $p < .001$, 95% CI = [0.84, 1.49]. Unexpectedly, we also found a significant conversation \times measurement phase interaction effect, $b = -0.62$, $SE = 0.23$, $t(296.00) = -2.63$, $p = .009$, 95% CI = [-1.08, -0.16]: Participants expected to feel equally happy after their shallow and deep conversations, $b = -0.07$, $SE = 0.17$, $t(296.00) = -0.40$, $p = .688$, 95% CI = [-0.39, 0.26], yet felt significantly happier after their deep conversations, $b = 0.55$, $SE = 0.17$, $t(296.00) = 3.31$, $p = .001$, 95% CI = [0.23, 0.87].

Mediational analyses. We found some evidence that underestimating the other person's care explained why participants expected their conversations to unfold less positively than they did. For the shallow conversations, partner care significantly mediated the effect of measurement phase (expectations vs. experiences) on connectedness, $b = -0.69$, 95% CI = [-1.17, -0.32], and happiness, $b = -0.37$, 95% CI = [-0.77, -0.05], but not awkwardness, $b = 0.07$, 95% CI = [-0.45, 0.52]. For the deep conversations, partner care significantly mediated the effect of measurement phase on awkwardness, $b = 0.73$, 95% CI = [0.05, 1.50], and connectedness, $b = -1.02$, 95% CI = [-1.65, -0.52], but not happiness, $b = -0.30$, 95% CI = [-0.73, 0.22].

Desired conversation intimacy. Replicating the previous experiments, participants reported that their typical conversations with strangers ($M = 3.55$, $SD = 2.33$) were significantly less intimate than they would ideally prefer ($M = 5.45$, $SD = 2.52$), *paired* $t(119) = -9.03$, $p < .001$, 95% CI_{difference} = [-2.32, -1.48], $d = -0.78$.

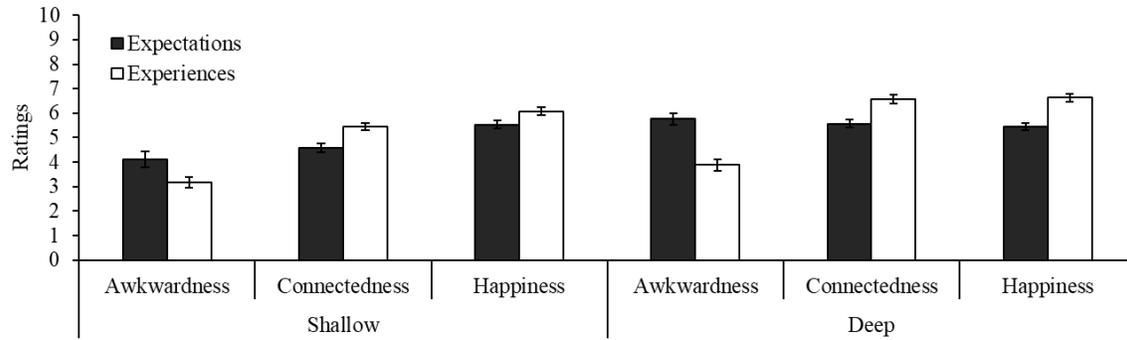


Figure 6. Mean awkwardness, connectedness, and happiness across conversation conditions (shallow vs. deep) and measurement phase (expectations vs. experiences) in Experiment 6a.

Error bars $\pm 1 SE$.

Results (Experiment 6b)

The strength-of-bond and liking items were highly correlated in both the shallow ($\alpha = .77$) and deep ($\alpha = .78$) conditions, as were the three friendship interest items ($\alpha_{\text{shallow}} = .92$, $\alpha_{\text{deep}} = .91$). We collapsed these items to form connectedness and friendship interest scales, respectively.

Manipulation check. The intimacy manipulation was effective: Participants reported that their deep conversations ($M = 6.78$, $SD = 2.06$) were significantly more intimate than their shallow conversations ($M = 2.19$, $SD = 2.17$), $\text{paired } t(119) = -16.22$, $p < .001$, $95\% \text{ CI}_{\text{difference}} = [-5.15, -4.03]$, $d = -2.17$.

Preference measures. Before engaging in the conversations, participants expected to prefer the shallow conversation relative to the deep conversation ($M = -1.56$, $SD = 3.25$), $\text{one-sample } t(119) = -5.26$, $p < .001$, $95\% \text{ CI} = [-2.15, -0.97]$, $d = -0.48$. After the conversations, however, participants reported preferring the deep conversation ($M = 0.88$, $SD = 3.26$), $\text{one-sample } t(119) = 2.94$, $p = .004$, $95\% \text{ CI} = [0.29, 1.46]$, $d = 0.27$. As a result, participants

significantly underestimated the extent to which they would prefer the deep conversation, *paired* $t(119) = -6.77, p < .001, 95\% CI_{\text{difference}} = [-3.14, -1.72], d = -0.75$.

After engaging in both conversations, participants also reported which questions they expected to prefer discussing in a future study session. Participants did not show a statistically significant preference for one conversation type over another ($M = 0.43, SD = 3.40$), *one-sample* $t(119) = 1.39, p = .166, 95\% CI = [-0.18, 1.05], d = 0.13$. This means that participants' reported preferences shifted significantly after having their conversations: Whereas participants expected to prefer the shallow conversation at the start of the study session, their predicted preferences shifted significantly in the direction of the deep conversation questions after actually engaging in both conversations, *paired* $t(119) = -6.16, p < .001, 95\% CI_{\text{difference}} = [-2.63, -1.35], d = -0.60$. This result suggests participants learned something during their conversations that they then used to update their expectations. Indeed, a post-hoc analysis indicates that participants' reported preferences before their conversations were only modestly correlated with their experienced preferences after their conversations, $r = .27, t(118) = 3.03, p = .003, 95\% CI = [.09, .43]$. In contrast, their experienced preferences were very strongly correlated with their preferences for a future conversation, $r = .73, t(118) = 11.77, p < .001, 95\% CI = [.64, .81]$. These two correlations differ significantly from each other, $Z = -6.13, p < .001$. Participants learned that the relatively deep conversation was more positive than the relatively shallow conversation, and used their direct experience to update their expectations for the future. Misunderstanding the outcomes of relatively deep conversation could create a barrier to engaging in them more often in everyday life, thereby keeping people from having the very experiences that would allow them to learn that their expectations could be miscalibrated (see also Epley & Schroeder, 2014, Experiment 4).

Experience measures. Participants preferred the deep conversation because their experiences were more positive overall during the deep conversation than the shallow conversation. We constructed mixed linear models with fixed-effects terms for conversation (shallow vs. deep) and random-intercept terms for the participant and the partner, separately for each measure. Participants cared significantly more about their own responses during the deep conversation than the shallow conversation, $b = -1.84$, $SE = 0.29$, $t(235.00) = -6.43$, $p < .001$, 95% CI = [-2.40, -1.28], and perceived the other person to care more about one's own responses during the deep conversation, $b = -1.37$, $SE = 0.28$, $t(238.00) = -4.90$, $p < .001$, 95% CI = [-1.91, -0.82]. Participants reported that their deep conversations felt only marginally more awkward than their shallow conversations, $b = -0.63$, $SE = 0.35$, $t(232.81) = -1.82$, $p = .070$, 95% CI = [-1.31, 0.05], but that they felt significantly more connected to their partner, $b = -1.10$, $SE = 0.25$, $t(238.00) = -4.33$, $p < .001$, 95% CI = [-1.59, -0.60], and more interested in being friends with the person, $b = -0.94$, $SE = 0.27$, $t(238.00) = -3.42$, $p < .001$, 95% CI = [-1.48, -0.40], after the deep conversation. Happiness did not differ significantly between the deep and shallow conversations, $b = -0.28$, $SE = 0.21$, $t(238.00) = -1.34$, $p = .182$, 95% CI = [-0.70, 0.13] (see Figure 7).

Desired conversation intimacy. Replicating the previous experiments, participants reported that their typical conversations with strangers ($M = 2.79$, $SD = 2.26$) were significantly less intimate than they would ideally prefer ($M = 5.54$, $SD = 2.65$), *paired* $t(119) = -11.11$, $p < .001$, 95% CI_{difference} = [-3.24, -2.26], $d = -1.11$.

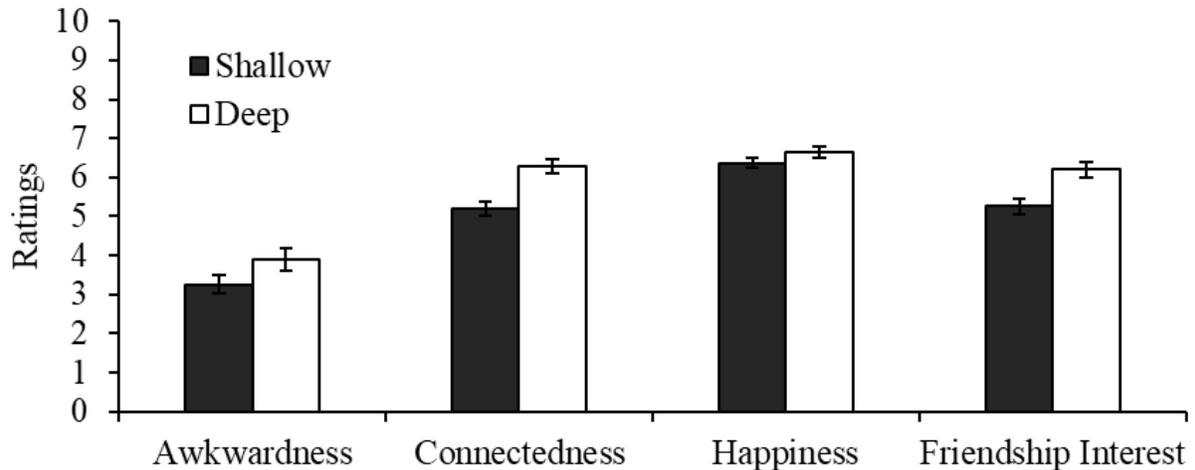


Figure 7. Mean awkwardness, connectedness, and happiness experiences across shallow and deep conversations in Experiment 6b. Error bars $\pm 1 SE$.

Discussion

Consistent with Experiments 1a-5, participants again generally underestimated how positive their conversations would be, and also overestimated how awkward deep conversations would be relative to shallow conversations. In contrast to some of the preceding experiments, participants also felt significantly more connected to their deep conversation partner than their shallow conversation partner, suggesting that people establish stronger connections through deep conversation—yet may be better able to report the relative strength of this connection when they compare two conversations directly.

More important, Experiment 6b provided direct measures of participants' preferences for relatively deep versus shallow conversations. Participants expected to prefer the shallow conversation more than the deep conversation before having the conversations, but consistent with our hypotheses, they reported actually preferring the deep conversation after having both of them. Not only does having a deep conversation with another person seem to be a surprisingly positive experience, it seems to be *more* positive than having a shallow conversation. As a result

of their direct experience, participants also appeared to learn that they would prefer a relatively deep conversation in the future more than their initial expectations suggested. This result highlights how miscalibrated expectations could create a barrier to engaging in deeper and more meaningful conversations, thereby encouraging shallower conversations than might be optimal for one's own relationships.

Experiments 7a and 7b test the final step in our theorizing. Specifically, we predict that directly manipulating the psychological barrier created by miscalibrated expectations would encourage people to have deeper conversations than they might otherwise. Because the preceding experiments suggest that people undervalue deep conversations at least partly because they underestimate how much others will care, and be interested in, the conversation's content, we manipulated how much participants would expect their conversation partner to care and then measured the intimacy of questions they were interested in discussing with their partner. We did so in Experiment 7a by manipulating whether people imagined talking with someone they knew to be very caring or a person they knew to be indifferent, and in Experiment 7b by either informing or not informing participants that their beliefs about others' care tend to be systematically miscalibrated. We expected participants to select deeper conversational topics to discuss when they imagined talking with a relatively caring person in Experiment 7a, and when they learned that people tend to underestimate how much others care in conversation (versus overestimate, or no information). If underestimating the extent to which others care about the content of conversation creates a barrier to deeper conversations, then removing it either by imagining conversation with a caring other or by calibrating participants' expectations should encourage deeper conversation. Note that this would occur only if participants were personally

interested in having deeper conversations more often in daily life, as participants in Experiments 1a, 1b, 1c, 4a, 4b, 6a, and 6b reported.

Experiments 7a-7b: Anticipated Indifference as a Barrier to Deep Talk

Pre-test

Before conducting the experiments, we pre-tested 20 discussion questions for perceived intimacy, some of which were used in Experiments 1a-1c, 2, 4a-4b, and 6a-b (see Supplemental Material for full details). The four perceived intimacy items were highly correlated ($\alpha = .96$) and so we combined them to form an intimacy scale. We designated the 10 questions with the highest average intimacy ratings as deep questions and the 10 questions with the lowest average intimacy ratings as shallow questions (see Appendix).

Method

Participants. We recruited 100 participants from Amazon's Mechanical Turk ($N = 93$ after exclusions; $M_{\text{age}} = 35.30$; $SD_{\text{age}} = 9.99$; 36.56% female; 78.49% Caucasian) to complete Experiment 7a in exchange for \$1.75. We excluded 7 participants from analyses because they failed the attention check described below.

We recruited 160 participants from Amazon's Mechanical Turk ($N = 130$ after exclusions; $M_{\text{age}} = 35.92$; $SD_{\text{age}} = 12.36$; 46.15% female; 75.38% Caucasian) to complete Experiment 7b in exchange for \$1.20. We excluded 30 of these participants because they failed one or more attention of the checks described below.

Procedure. Participants in Experiment 7a imagined visiting a “social interaction” research lab where they would speak with another study participant they had never met before in a discussion called the “sharing game.” Participants were then randomly assigned to the caring or uncaring condition. Participants in the “caring” condition read, “Before the session begins,

suppose you see the other person in the waiting room. You've seen this person around and you have the impression that this person is very sociable, caring, and considerate of others. Although you don't know this person, you feel pretty confident in your judgment.” Participants in the “uncaring” condition read, “Before the session begins, suppose you see the other person in the waiting room. You've seen this person around and you have the impression that this person is rather indifferent toward others and isn't very caring or considerate. Although you don't know this person, you feel pretty confident in your judgment.”

Participants in both the caring and uncaring conditions then read the 20 possible discussion questions and selected 5 they would most like to ask and answer while speaking with the other person. The questions were presented in random order and not labeled in any way. After selecting 5 questions, participants then viewed each of the 20 questions separately and for each question reported how much they wanted to discuss the question with the other person (0 = *not at all*; 10 = *quite a bit*), how much they would care about their own response to the question (0 = *not at all*; 10 = *quite a bit*), and how much they believed the other person would care about the participant's response to the question (0 = *not at all*; 10 = *quite a bit*). After evaluating the questions, participants completed an attention check by reporting whether the other participant was described as sociable, caring, and considerate or as indifferent and not very caring or considerate. Finally, participants reported demographic information and received payment.

The procedure for Experiment 7b was identical to Experiment 7a with three exceptions. First, we designed manipulations that either did or did not inform participants that people's expectations tend to be systematically miscalibrated. Participants in the *underestimation* condition read, “In these experiments, we find that people tend to UNDERESTIMATE how much strangers will care about each other's responses to these questions. That is, strangers tend

to be MORE concerned and interested in each other's responses than people expect." Participants in the *overestimation* condition read, "In these experiments, we find that people tend to OVERESTIMATE how much strangers will care about each other's responses to these questions. That is, strangers tend to be LESS concerned and interested in each other's responses than people expect." Participants in the *control* condition were not told anything about the extent to which others' expectations tend to be calibrated. Second, after selecting 5 out of the 20 questions, participants in this experiment reported how interested they would be in asking and answering each of the 20 questions with the other person (0 = *not at all*; 10 = *extremely*), but did not predict the degree of care for either themselves or their conversation partner. Third, we tailored the attention checks to the current procedure by asking participants to report whether they imagined speaking with a friend or stranger, and to report what they had been told about the results of our previous research experiments.

Results

We pre-registered analyses of the number of deep and shallow questions participants selected across conditions in both Experiments 7a and 7b, but also report analyses of the average intimacy level of the questions selected.

As predicted, participants in the caring condition of Experiment 7a ($M = 3.09$, $SD = 1.44$) selected significantly more deep questions than did participants in the uncaring condition ($M = 2.02$, $SD = 1.73$), $t(91) = 3.23$, $p = .002$, 95% $CI_{\text{difference}} = [0.41, 1.72]$, $d = 0.67$. Participants in the caring condition also selected questions higher in average intimacy ($M = 4.96$, $SD = 1.22$) than did participants in the uncaring condition, ($M = 4.24$, $SD = 1.48$), $t(91) = 2.53$, $p = .013$, 95% $CI_{\text{difference}} = [0.15, 1.27]$, $d = 0.53$. These differences in choice could arise either because participants in the caring condition have a stronger desire to discuss deep questions or because

they have a weaker desire to discuss shallow ones, compared to participants in the uncaring condition. We therefore computed each participant's mean desire to discuss the 10 shallow questions and the 10 deep questions, separately. Consistent with our predictions, participants in the caring condition were particularly interested in discussing deep questions. A 2 (partner: caring, uncaring) \times 2 (question type: shallow, deep) ANOVA with repeated measures on the second factor yielded a significant partner \times question type interaction, $F(1, 91) = 9.52, p = .003, \eta_p^2 = .09$ (see Figure 8): Participants in the caring and uncaring conditions did not differ in their desire to discuss shallow questions, $t(91) = -0.70, p = .486, 95\% \text{ CI}_{\text{difference}} = [-1.07, 0.51], d = -0.14$, but those in the caring condition were significantly more interested in discussing deep questions, $t(91) = 3.30, p = .001, 95\% \text{ CI}_{\text{difference}} = [0.54, 2.16], d = 0.68$. Underestimating others' care may thus create a barrier to having deeper and more intimate conversations.

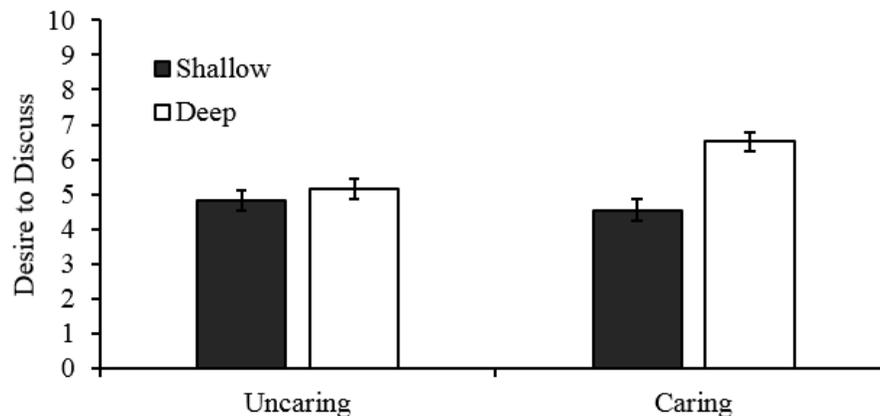


Figure 8. Mean desire to discuss across partner (uncaring vs. caring) and question type (shallow vs. deep) in Experiment 7a. Error bars $\pm 1 \text{ SE}$.

In Experiment 7b, a one-way ANOVA indicated that the number of deep questions selected varied significantly across care conditions, $F(2, 127) = 3.71, p = .027, \eta_p^2 = .06$. As predicted, participants in the underestimation condition ($M = 2.41, SD = 1.12$) selected significantly more deep questions than participants in the overestimation condition ($M = 1.69, SD = 1.26$), $t(127) = -2.68, p = .008, 95\% CI_{\text{difference}} = [-1.26, -0.19], d = -0.57$. The number of deep questions selected in the control condition fell in between ($M = 2.15, SD = 1.44$), differing neither from the underestimation condition, $t(127) = -0.95, p = .344, 95\% CI_{\text{difference}} = [-0.83, 0.29], d = -0.21$, nor the overestimation condition, $t(127) = -1.69, p = .094, 95\% CI_{\text{difference}} = [-1.00, 0.08], d = -0.36$. Likewise, the mean intimacy of selected questions varied by condition, $F(2, 127) = 3.40, p = .036, \eta_p^2 = .05$, with participants in the underestimation condition ($M = 4.53, SD = 1.02$) selecting questions of higher average intimacy than participants in the overestimation condition ($M = 3.92, SD = 1.10$), $t(127) = -2.60, p = .010, 95\% CI_{\text{difference}} = [-1.06, -0.14], d = -0.55$. The mean intimacy of selected questions in the control condition fell in between ($M = 4.16, SD = 1.14$), and did not differ significantly from either the underestimation condition, $t(127) = -1.51, p = .133, 95\% CI_{\text{difference}} = [-0.84, 0.11], d = -0.33$, or the overestimation condition, $t(127) = 1.03, p = .304, 95\% CI_{\text{difference}} = [-0.22, 0.70], d = 0.22$.

As with Experiment 7a, we tested whether these differences arose from differences in interest in discussing deep questions, shallow questions, or possibly both. We did so by comparing participants' reported desire to discuss the 10 deep and 10 shallow questions, on average, between the underestimation and overestimation conditions. A 2 (care: underestimation, overestimation) \times 2 (question type: shallow, deep) ANOVA with repeated measures on the second factor and mean interest as the dependent measure yielded a significant main effect of care, $F(1, 87) = 5.89, p = .017, \eta_p^2 = .06$, indicating that participants in the underestimation

condition were more interested in discussing the questions overall, and a significant main effect of question type, $F(1, 87) = 4.62, p = .034, \eta_p^2 = .05$, indicating that participants in both conditions were more interested in discussing deep versus shallow questions. The care \times question type interaction effect was non-significant, $F(1, 87) = 0.69, p = .410, \eta_p^2 = .01$. However, planned contrasts provide some evidence that participants in the underestimation condition may have been somewhat more interested in discussing deep questions: Participants in the underestimation condition reported only marginally greater interest in discussing shallow questions than participants in the overestimation condition, $t(127) = -1.68, p = .095, 95\% \text{ CI}_{\text{difference}} = [-1.65, 0.13], d = -0.36$, but reported significantly greater interest in discussing deep questions, $t(127) = -2.68, p = .008, 95\% \text{ CI}_{\text{difference}} = [-2.04, -0.31], d = -0.57$ (see Figure 9).

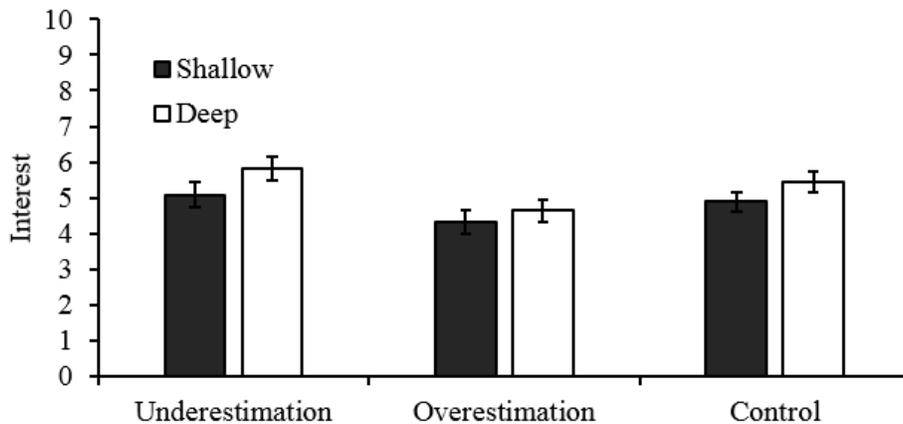


Figure 9. Mean interest across care (overestimation vs. underestimation vs. control) and question type (shallow vs. deep) in Experiment 7b. Error bars $\pm 1 \text{ SE}$.

Discussion

Underestimating the extent to which others care about the content of deep conversation can create miscalibrated expectations about the outcomes of these conversations, and also create

a barrier to engaging in these conversations in the first place. Experiments 7a and 7b suggest that manipulating this barrier directly may influence people's preferences for deeper conversations with strangers. Participants in Experiment 7a chose deeper questions when they expected their conversation partner to be caring rather than indifferent. Participants in Experiment 7b likewise chose deeper questions when they were told that people tend to underestimate strangers' care than when they were instead told that people tend to overestimate strangers' care. It is important to note that although the control condition in Experiment 7b was different from neither the overestimation nor underestimation condition, it was at least descriptively more similar to the underestimation condition. We are not sure how to interpret this result because we did not obtain a manipulation check to test how much this particular experimental manipulation actually affected participants' expectations. It could be that it is simply easier to convince people that others care less than expected about a deep conversation than to convince them that others care more. These particular results therefore confirm that manipulating perceptions of others' care can affect preferences for deep talk, but they cannot at this point confirm whether weakening or strengthening that barrier is likely to have a larger impact on people's social preferences.

General Discussion

Human beings are a deeply social species, whose wellbeing and health are tightly tied to the quality of their social relationships (Diener & Seligman, 2002; Holt-Lunstad, 2018; Sun, Harris, & Vazire, 2019). It is therefore not surprising that most people have strong desires to strengthen and maintain meaningful relationships (Baumeister & Leary, 1993; Ryan & Deci, 2000), actively combatting feelings of loneliness and isolation by seeking social connection with others (Maner, DeWall, Baumeister, & Schaller, 2007). Disclosing meaningful and intimate information with others in conversation is one way to strengthen social relationships (Aron, et

al., 1997; Collins & Miller, 1994), and so it is also not surprising that participants in our experiments reported wanting to have more deep and meaningful conversations with others more often in their daily lives. If people want deeper conversations with others, then why aren't they having them?

Our data suggest that underestimating others' deeply social nature—assuming that others will be more indifferent and uncaring in conversation than they actually are—could help to explain why conversations in daily life are shallower than people might prefer. Our participants consistently expected their conversations to be more awkward, and to lead to weaker connections and less happiness, than they actually did regardless of whether participants generated the topics of conversation (Experiments 3, 5) or not (Experiments 1a, 1b, 1c, 2, 4a, 4b, 6a, 6b). Participants also consistently overestimated how awkward deep conversations would be more so than shallow conversations (Experiments 2-6a), but were more calibrated for conversations with close others whose care and interest were more clearly known (Experiment 5). Finally, participants chose deeper questions when they expected their conversation partner to be relatively more caring (Experiments 7a-b), providing the clearest evidence that miscalibrated inferences about others' sociality may act as a psychological barrier to deeper conversations.

Limitations

Although participants' tendency to underestimate the positive outcomes of deep conversation was robust across important methodological variations, there are at least three potential limitations on the generalizability of our results.

Our first limitation on generalizability is that all of our experiments manipulated intimacy in experimental settings in which both people in the conversation knew they were going to have a conversation as part of an experiment. This is obviously not the case in naturalistic

conversations in daily life in which people are likely to be more uncertain about whether another person is likely to reciprocate deeper and more intimate disclosures, and also must infer why another person is asking relatively deeper or shallower conversation topics. Asking another person to share the last time they cried in front of someone else might be received much differently when you raise the question to the person standing next to you in the supermarket line than it was in our experiments. Of course, any differences between our experimental procedures and more naturalistic circumstances are also likely to affect people's expectations of the interaction in a way that could still maintain the critical gap we observed between the expected and actual outcomes of more naturalistic conversations. The deeper question one might consider asking in the supermarket line might still yield a better conversation than expected.

In fact, we think people might be even more likely to underestimate the positive outcomes of deep conversations in naturalistic settings outside the lab. Previous research in both the United States and in the United Kingdom indicates that people may underestimate strangers' willingness to engage in conversation in naturalistic field experiments on trains, busses, and cabs, which can partly then explain people's tendency to underestimate how much they would enjoy having a conversation with a stranger (Epley & Schroeder, 2014; Schroeder, Lyons, & Epley, 2021; see also Sandstrom & Boothby, 2021). Our experimental procedures eliminate this concern by instructing participants to speak with another study participant who they know will engage them in conversation. Our experiments may therefore provide a more conservative test of whether people underestimate the value of deep conversations compared to more naturally occurring conversations. In addition, starting a meaningful conversation signals one's trust and interest in the listener. This signal may be more influential when the listener cannot attribute the depth of conversation to an experimental instruction. In daily life, listeners may be even more

concerned about intimate details that another person freely shares with them. Nevertheless, it could also be that listeners form more negative inferences about another person's interest or intent when they raise deeper conversation topics in daily life, such as inferring that another person is interested in a romantic relationship or is potentially trying to manipulate or take advantage of a listener in some way. It is possible these inferences are not anticipated by those trying to start a conversation, in which case the gap between expected and actual outcomes of deeper conversations might vary across contexts.

Future research could examine the generalizability of our results in more naturalistic contexts by manipulating whether both participants are aware of the conversation instructions or not. To the extent that both participants reveal meaningful information about themselves during the conversation, our theory predicts that deep conversations should lead to similar, if not even stronger, connections when one person is unaware of the conversation instructions. The generalizability of our effects could also be tested using diary methods in which people are asked to engage in relatively deeper or shallower conversations over the course of a day and then report their experiences at the end of the day, or simply by asking people to engage in a deeper conversation when they have a natural opportunity to do so in daily life. Existing research using this methodology has found that people report more positive affect when asked to act extraverted than when asked to act introverted (Margolis & Lyubomirsky, 2020), when asked to engage in conversation on a train or bus compared to remaining silent (Epley & Schroeder, 2014; Schroeder, Lyons, & Epley, 2021), or simply when asked to try to make a connection to a coffee shop barista compared to being as efficient as possible (Sandstrom & Dunn, 2014a; see also Gunaydin, Oztekin, Karabulut, & Salman-Engin, 2021). These existing results, showing positive

affect resulting from simply engaging with others in naturalistic settings, give us reason to believe that engaging more deeply in these same settings would enhance positive affect as well.

A second potential limitation on generalizability is that our experiments investigated only short-term outcomes of a single relatively shallow and/or deep conversation. There could be longer-term consequences, either more positive or negative, of these conversations that we did not measure. For instance, participants in Experiment 2 felt somewhat lonelier after having a relatively deep conversation than after a shallower conversation. Although we only measured loneliness in one experiment, and hence do not know how robust this result is, it could be that a single conversation highlights dissatisfying features of a person's other relationships, such as other relationships being shallower than one might desire, that affect a person's wellbeing in a more complicated fashion beyond the single interaction we examined. It could also be that people feel worse over time after a very satisfying conversation with a stranger after recognizing that they are not likely to have another conversation with that person again. Again, field experiments involving diary methods would be an appropriate way of examining longer-term outcomes of having relatively more or less deep conversations in everyday life.

A final potential limitation on the generalizability of our results comes from conducting all of our experiments within in a relatively narrow time period within the United States. Although our samples are relatively diverse, including university students, master's degree students from around the world, and corporate employees and executives in the United States, it is reasonable to wonder how our effects may—or may not—vary across cultures. Cultures may vary in their openness to strangers, with individualistic cultures generally being more accepting and open to strangers and collectivistic cultures prioritizing ingroup relationships. If cultural differences exist in the willingness to engage in deep and meaningful conversations with

strangers, then they could arise either from cultural differences in expectations or from cultural differences in actual experiences. Given the fundamental importance of social relationships across humans' evolutionary history (Dunbar, 1998; von Hippel, 2018), we would predict that expectations about others' care and interest are more likely to vary across cultures than the actual experience of deep and meaningful conversation. Cultures could vary in the relative frequency of deep versus shallow conversations because they vary in their social expectations rather than in their social experiences. Note that the social expectations we have identified are the very kind that would create social avoidance and hence keep people from learning about the actual consequences of relatively shallow versus deep conversations. In Experiment 6b, participants reported a stronger preference for deep conversation after having both a shallow and a deep conversation, preferences that our data suggest would be more closely aligned with optimizing people's subjective experience and the quality of their relationships. Similarly, individuals within a given culture might vary in their social expectations, such that those who are highly socially avoidant (Strachman & Gable, 2006), high in rejection sensitivity (Pietrzak, Downey, & Ayduk, 2005), or insecurely attached (Keelan, Dion, & Dion, 1998) might also be especially pessimistic about deeper conversations, and hence avoid them more often in daily life.

Miscalibrated expectations may therefore come partly from avoiding the experiences that would otherwise calibrate a person's expectations.

Unanswered Questions

We believe our current results make important theoretical contributions to the literatures on both self-disclosure and wellbeing. Decisions to open up to another person or remain more guarded are based in part on how people expect that their interaction partner will respond (Afifi & Steuber, 2009; Kardas, Kumar, & Epley, 2021; Omarzu, 2000; Ruan, Reis, Clark, Hirsch, &

Bink, 2020), yet psychologists have primarily examined the causes (Altman & Taylor, 1973; Berg & Clark, 1986; Cline, 1989) and consequences (Collins & Miller, 1994; Kelly & McKillop, 1996) of self-disclosure separately from one another. Our paper highlights the importance of combining these two streams of research. Understanding how miscalibrated inferences about others' social interest and care guide decisions about the content of conversation can help to explain why people may be missing out on opportunities to establish stronger connections with others.

We believe our research also highlights the importance of measuring the accuracy of the causes underlying self-disclosure, in this case the calibration between expectations of others' reactions in conversation and their actual reactions. Doing so opens up new questions for theoretical inquiry, especially whether people disclose *enough* deep and intimate content in their interactions to maximize the quality of relationships. Experiments 1a-5 do not answer this question definitively. Although participants tended to underestimate how positive deep conversation would be overall, the deep conversations were indeed somewhat more awkward than the shallow conversations in actual experience. Instead of guessing about how people themselves might integrate these unique aspects of their experience into an overall preference for conversations, we enabled participants to compare these two conversations against each other directly in Experiments 6a and 6b by instructing them to have both a deep and a shallow conversation. These experiments provided the clearest evidence that people uniquely underestimate the positive outcomes of deep conversation, that deep conversations indeed create more positive experiences than shallow conversations overall, and that people report preferring the deeper conversation afterwards to a greater extent than they expected beforehand. These

results suggest that miscalibrated expectations may leave people being overly reluctant to discuss deep and intimate content in their everyday conversations with strangers.

Although participants significantly preferred relatively deeper conversations to shallower conversations in Experiments 6a and 6b, it is worth noting that participants across our experiments also tended to report positive experiences after relatively shallow conversations, and tended to report more positive (and slightly less negative) experiences in these conversations than they anticipated as well. These results are consistent with prior research suggesting a more general phenomenon in which conversations with strangers tend to be surprisingly pleasant (Epley & Schroeder, 2014; Sandstrom & Boothby, 2021). Indeed, we observed somewhat inconsistent differences in experienced connectedness, happiness, and enjoyment across the experiments in which participants experienced only a relatively deep or shallow conversation. For instance, participants in the deep condition felt significantly more connected to their conversation partner than did participants in the shallow or control condition in Experiments 3 ($p = .022$, $d = 0.47$) and 4a ($p = .016$, $d = 0.48$), but did not differ significantly in Experiments 2 ($p = .135$, $d = 0.32$) and 4b ($p = .970$, $d = -0.01$). Participants in the deep condition also felt marginally happier about their conversation than did participants in the shallow or control condition in Experiment 3 ($p = .088$, $d = 0.34$), but did not report feeling significantly happier in Experiments 4a ($p = .962$, $d = -0.01$), 4b ($p = .309$, $d = -0.20$), or in Experiment 5 among participants who spoke with either distant strangers ($p = .190$, $d = 0.26$) or close others ($p = .160$, $d = 0.28$). These somewhat inconsistent results are surprising because previous literature suggests that intimate self-disclosure should lead to more positive conversations than impersonal disclosures (Aron et al., 1997; Collins & Miller, 1994), and our manipulations were designed precisely to encourage more intimate disclosures in the deep conditions.

Inconsistency in the connection value of deep conversation may come partly from the unique nature of our experimental exchanges compared to the existing literature. Many prior experiments have more carefully constrained the breadth and depth of self-disclosure through impression-formation tasks in which participants read about or observe others' interactions (e.g., Berg & Archer, 1980; Bradac, Hosman, & Tardy, 1978; Chaikin & Derlega, 1974), or by using written essays and self-descriptions (Brewer & Mittelman, 1980; Rubin, 1975). Even procedures involving live, spoken interactions have often required that participants deliver uninterrupted verbal self-descriptions (Chaikin, Derlega, Bayma, & Shaw, 1975; Jones & Archer, 1976) or engage in structured question-answer sequences (Sprecher, Treger, Wondra, Hilaire, & Wallpe, 2013). Each of these methodologies carefully controls the amount of self-disclosure and thereby tests the impact of self-disclosure on liking and connectedness. However, doing so also reduces ecological validity by constraining the natural flow of back-and-forth conversation. In contrast, participants in Experiments 1a-5 engaged in back-and-forth conversations and were not explicitly forbidden from discussing content beyond the scope of the discussion questions. Participants in Experiments 6a and 6b were encouraged not to stray from the questions we asked them to discuss, but their conversations were not monitored or constrained beyond that instruction. Overall, our experiments may better approximate conversation in everyday life, but they also sacrifice the tight experimental control that would come from more scripted or one-sided exchanges.

Rather than call into question existing literature on self-disclosure, we believe that our findings may suggest an important moderator to explain when, and why, intimacy manipulations influence positive outcomes such as liking and connectedness (Collins & Miller, 1994). Intimacy manipulations may have greater impact in static exchanges such as impression-formation tasks,

written exchanges, and structured question-answer sequences than in dynamic exchanges involving back-and-forth conversation. Indeed, the small number of prior experiments that have manipulated intimacy in less structured interactions have produced mixed findings. Some experiments have produced differences in experienced connectedness (Aron et al., 1997; Shearer, 2017), while others have produced mixed evidence across multiple experiments (Kashdan, McKnight, Fincham, & Rose, 2011) or no differences (Besst, 2016; Kashdan & Roberts, 2006). Because conversation naturally becomes more intimate as people continue talking (Taylor, 1968), even relatively superficial questions may promote liking and connectedness by acting as a gateway to somewhat deeper conversation, thereby promoting relatively positive outcomes regardless of the initial depth of the conversation.

As already mentioned, anecdotal reports from the live debriefing of Experiment 4b support this possibility, with participants describing how surprisingly deep and meaningful their ostensibly “shallow” conversations became as they unfolded. Some results from Experiment 5 are at least consistent with this possibility of increasing intimacy even in what we intended to be relatively shallow conversations. In the control conditions, participants who spoke with distant strangers perceived their conversations ($M = 4.21$, $SD = 2.22$) to be significantly more intimate than the questions they discussed ($M = 2.50$, $SD = 2.40$), $paired\ t(49) = -4.98$, $p < .001$, 95% $CI_{\text{difference}} = [-2.40, -1.02]$, $d = -0.70$, and those who spoke with close others likewise perceived their conversations ($M = 5.03$, $SD = 2.46$) to be significantly more intimate than the questions they discussed ($M = 3.70$, $SD = 2.65$), $paired\ t(49) = -3.69$, $p < .001$, 95% $CI_{\text{difference}} = [-2.05, -0.61]$, $d = -0.52$. In contrast, these patterns weakened or reversed in the deep conditions, where participants who spoke with distant strangers perceived their conversations ($M = 6.41$, $SD = 1.67$) to be *less* intimate than the questions they discussed ($M = 7.04$, $SD = 1.58$), $paired\ t(49) =$

2.09, $p = .042$, 95% $CI_{\text{difference}} = [0.02, 1.24]$, $d = 0.30$, and those who spoke with close others likewise perceived their conversations ($M = 7.49$, $SD = 1.86$) to be no more intimate than the questions they discussed ($M = 7.78$, $SD = 1.43$), $paired\ t(49) = 1.22$, $p = .230$, 95% $CI_{\text{difference}} = [-0.19, 0.77]$, $d = 0.17$.

These results raise two interesting topics for future research. First, relatively shallow questions could act as a gateway to deeper conversation, potentially helping to explain why our deep conversation conditions did not consistently create significantly stronger experiences of connectedness, happiness, and enjoyment in our experiments than our comparison conditions. Second, our experiments attempted to hold the depth of conversation topics constant across conditions, but self-disclosure may naturally change over the course of naturalistic conversations. People may expect that conversations are better when they become increasingly intimate over the course of the conversation, and hence start conversations more shallowly than they hope to end them (Aron et al., 1997). It would be interesting and important to understand how people's expectations about changes in intimacy across conversations are aligned with their actual experiences.

If unscripted conversations do indeed tend to change over the course of the conversation in ways that people do not fully anticipate, then this could also help explain one other unexpected result: In Experiments 4a and 4b, participants in the shallow condition expected to feel significantly less connected to each other than participants in the deep condition, yet both groups actually felt strongly connected after speaking. As a result, participants in the shallow condition underestimated how connected they would feel *more* than did participants in the deep condition. Natural conversation may transform relatively shallow topics into relatively deeper content as the conversation continues. If people overlook this property of conversation and

instead focus myopically on the topics that they plan to discuss—in a similar way that people expect their emotional reactions to major life events to be determined largely by the event in question and not by other future events or by their ability to make sense of the event (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000)—then they could underestimate positive outcomes more for shallow questions than for deep questions due to exaggerated differences in their expectations. This hypothesis predicts that people should underestimate the positive outcomes of conversation even more for *trivial* topics that initially require virtually no self-disclosure. In contrast, people should have more accurate expectations of more highly structured question-answer sequences that constrain the natural dynamics of back-and-forth conversation and thus keep conversations from becoming increasingly intimate. We believe this increasing intimacy hypothesis about unconstrained conversation warrants further investigation because it may help to explain a range of findings indicating that people underestimate how much they will enjoy connecting with distant strangers (Dunn, Biesanz, Human, & Finn, 2007; Epley & Schroeder, 2014; Mallett, Wilson, & Gilbert, 2008).

Our findings also raise questions about other novel mechanisms guiding social behavior that may lead people to mismanage their relationships. One question emerges from the different patterns of results obtained on our different outcome measures: In most of our experiments, participants underestimated their connectedness and happiness, and their partner's care, equally for deep and shallow conversations yet overestimated their awkwardness more for deep conversations. Moreover, we found stronger evidence that underestimating others' care explained participants' miscalibrated expectations of connectedness and happiness than awkwardness. One explanation for these divergent findings concerns the expected ambivalence

of deep talk: Participants generally expected deep conversations to create more positive experiences of connectedness and happiness *and* more negative experiences of awkwardness than superficial conversations. Prior research suggests that as an event approaches in time, people may focus more narrowly on its perceived risks relative to its perceived rewards (Baumeister et al., 2001; Brown, 1948; Epstein & Fenz, 1962; Miller, 1944). This heightened sensitivity to immediate risks could cause people to expect immediate deep conversations to be particularly awkward, creating a reluctance to discuss deep topics in the midst of an ongoing conversation—even if they recognize in the abstract that they would prefer to have deeper conversations more often in their daily lives. Such a mechanism could lead to excess avoidance in a wide range of social situations because almost any decision to engage with others can lead to either positive outcomes of social connection and happiness or negative outcomes of awkwardness stemming from others' indifference or outright hostility. Future research could test this explanation by manipulating whether a deep conversation will take place immediately or in the distant future and measuring whether participants expect immediate deep conversations to be more negative than distant ones. Another possible explanation is that multiple classes of mechanisms contributed to the patterns of miscalibrated expectations observed in our experiments: Apart from expecting *others* to be less caring than they were, participants may have failed to appreciate how comfortable and positive *they* would personally feel about disclosing self-relevant information to another person. Self-disclosure may feel less awkward than people expect, independent of feedback received from the listener. If so, this finding would bridge the literatures on social judgment and affective forecasting, which have been studied separately but could operate jointly to create unnecessary barriers to self-disclosure (see also Leunissen, De Cremer, van Dijke, & Folmer, 2014).

Broader Implications

We document that people undervalue the positive consequences of conversations, especially deep conversations, at least partly because they underestimate how much strangers in conversation will be interested in the content of the conversation, and care about the intimate information being shared. We believe this is part of broader tendency to underestimate others' sociality, thereby creating a wide variety of psychological barriers to social engagement. People may underestimate how interested strangers on trains, buses, and cabs are in talking, thereby creating reluctance to start a conversation and increase one's wellbeing (Epley & Schroeder, 2014). Similarly, people underestimate how positive others will feel following an expression of gratitude (Kumar & Epley, 2018), a random act of kindness (Kumar & Epley, 2021), a compliment (Boothby & Bohns, 2020; Zhao & Epley, in press), a trusting disclosure of a negative secret (Kardas, Kumar, & Epley, 2021), and even a constructive confrontation in an established relationship (Dungan & Epley, 2021). People expect others to be more indifferent to these prosocial acts than others actually are, just as we found for deep conversation. There are several possible explanations for these effects. One is that people tend to dehumanize others by assuming that others have weaker mental capacities, including capacities for empathy and compassion, than they do themselves (Haslam, Bain, Douge, Lee, & Bastian, 2005; Waytz, Schroeder, & Epley, 2014). Another is that egocentric biases in judgment may lead people to undervalue the importance of expressing warmth in others' evaluations. Actors in a situation tend to focus relatively more on competency-related aspects of their actions—how well one might express intimate thoughts in a conversation or in writing—while observers or recipients of these actions tend to be focused on warmth-related aspects of the same actions (Fiske, Cuddy, & Glick, 2007; Wojciszke, 1994). If expectations of social interactions are guided by a heightened

attention to *how* they are connecting to others—such as by focusing on the content of relatively shallow or deep conversation topics—but experiences are guided more heavily by the warmth conveyed in those interactions—such as the trust conveyed by revealing intimate details in a deep conversation—then people could systematically underestimate the power of social motivations in others. We believe that understanding these deeper mechanisms is a critical topic for future research because it could explain a wide variety of seemingly related phenomena in social life.

Finally, we believe our findings have important practical implications. Social life affords many opportunities to engage, and to engage more deeply, with others around us. Our participants' expectations about deeper conversations were not woefully misguided, but they were reliably miscalibrated in a way that could keep people from engaging a little more deeply with others in their daily lives. These miscalibrated expectations could matter for two reasons. First, by creating a barrier to engaging with others, these expectations could be self-fulfilling because they keep people from obtaining data that could otherwise calibrate their expectations. If you think that a deep conversation is likely to be especially awkward, then you are unlikely to give yourself the chance to find out that you might be a little bit wrong. Only by engaging with others do people accurately understand the consequences of doing so. Second, strengthening social relationships is critical for wellbeing, meaning that a reluctance to engage more deeply with others may leave people being less social than would be optimal for their own wellbeing. Being willing to dig a little deeper than one might normally go in conversation brings the opportunity to create a stronger sense of connection with others, especially with strangers. The world that Schopenhauer described of prickly porcupines existing a moderate distance from each other describes a social life that *could be*, not the one that *has to be*.

References

- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470-474.
- Afifi, T., & Steuber, K. (2009). The revelation risk model (RRM): Factors that predict the revelation of secrets and the strategies used to reveal them. *Communication Monographs*, *76*, 144-176.
- Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.
- Ames, D., Maissen, L. B., & Brockner, J. (2012). The role of listening in interpersonal influence. *Journal of Research in Personality*, *46*, 345-349.
- 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.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323-370.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497-529.
- Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of Political Economy*, *101*, 385-409
- Berenson, K. R., Gyurak, A., Ayduk, Ö., Downey, G., Garner, M. J., Mogg, K., ... & Pine, D. S. (2009). Rejection sensitivity and disruption of attention by social threat cues. *Journal of Research in Personality*, *43*, 1064-1072.
- Berg, J. H., & Archer, R. L. (1980). Disclosure or concern: A second look at liking for the norm

- breaker. *Journal of Personality*, 48, 245-257.
- Berg, J. H., & Clark, M. S. (1986). Differences in social exchange between intimate and other relationships: Gradually evolving or quickly apparent? In *Friendship and social interaction* (pp. 101-128). Springer, New York, NY.
- Besst, S. (2016). *The Level of Relational Self-Construal Moderates the Relationship between Disclosure and Well-Being*. (Unpublished honors thesis). College of Saint Benedict and Saint John's University, Saint Joseph, USA.
- Boothby, E. J., & Bohns, V. K. (2020). Why a simple act of kindness is not as simple as it seems: Underestimating the positive impact of our compliments on others. *Personality and Social Psychology Bulletin*, 47, 826-840.
- Boothby, E. J., Clark, M. S., & Bargh, J. A. (2017). The invisibility cloak illusion: People (incorrectly) believe they observe others more than others observe them. *Journal of Personality and Social Psychology*, 112, 589-606.
- Bradac, J. J., Hosman, L. A., & Tardy, C. H. (1978). Reciprocal disclosures and language intensity: Attributional consequences. *Communications Monographs*, 45, 1-17.
- Brewer, M. B., & Mittelman, J. (1980). Effects of normative control of self-disclosure on reciprocity. *Journal of Personality*, 48, 89-102.
- Brown, J. S. (1948). Gradients of approach and avoidance responses and their relation to level of motivation. *Journal of Comparative and Physiological Psychology*, 41, 450-465.
- Bruck, A., Scholl, S. G., & Bless, H. (2018). Beautiful mess effect: Self–other differences in evaluation of showing vulnerability. *Journal of Personality and Social Psychology*, 115, 192-205.
- Chaikin, A. L., & Derlega, V. J. (1974). Variables affecting the appropriateness of self-

- disclosure. *Journal of Consulting and Clinical Psychology*, 42, 588-593.
- Chaikin, A. L., Derlega, V. J., Bayma, B., & Shaw, J. (1975). Neuroticism and disclosure reciprocity. *Journal of Consulting and Clinical Psychology*, 43, 13-19.
- Cislak, A., & Wojciszke, B. (2008). Agency and communion are inferred from actions serving interests of self or others. *European Journal of Social Psychology*, 38, 1103-1110.
- Cline, R. J. W. (1989). The politics of intimacy: Costs and benefits determining disclosure intimacy in male-female dyads. *Journal of Social and Personal Relationships*, 6, 5-20.
- Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, 116, 457-475.
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92, 631-648.
- Derlega, V. J., Winstead, B. A., & Greene, K. (2001). Self-disclosure and starting a close relationship. *Handbook of Relationship Beginnings*, 153-74.
- Diener, E., & Seligman, M. E. (2002). Very happy people. *Psychological Science*, 13, 81-84.
- Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, 6, 178-190.
- Dungan, J. A., & Epley, N. (2021). *Surprisingly good talk: Misunderstanding others creates a barrier to constructive confrontation*. Unpublished manuscript, University of Chicago.
- Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending money on others promotes happiness. *Science*, 319, 1687-1688.
- Dunn, E. W., Biesanz, J. C., Human, L. J., & Finn, S. (2007). Misunderstanding the affective consequences of everyday social interactions: The hidden benefits of putting one's best face forward. *Journal of Personality and Social Psychology*, 92, 990-1005.

- Epley, N., & Schroeder, J. (2014). Mistakenly seeking solitude. *Journal of Experimental Psychology: General*, *143*, 1980-1999.
- Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (5th ed., pp. 498–541). New York, NY: Wiley.
- Epstein, S., & Fenz, W. D. (1962). Theory and experiment on the measurement of approach-avoidance conflict. *The Journal of Abnormal and Social Psychology*, *64*, 97–112.
- Fishman, M. & Gardner, W. (2017). *Stranger danger? Removing barriers to social connection through meaningful conversation* (Unpublished honors thesis). Northwestern University, Chicago, USA.
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*, 77-83.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878-902.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21-38.
- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: a source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*, 617-638.
- Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504-528.
- Gunaydin, G., Oztekin, H., Karabulut, D. H., & Salman-Engin, S. (2021). Minimal social

- interactions with strangers predict greater subjective well-being. *Journal of Happiness Studies*, 22, 1839-1853.
- Hart, E., VanEpps, E., & Schweitzer, M. (2021). The (better than expected) consequences of asking sensitive questions. *Organizational Behavior and Human Decision Processes*, 162, 136-154.
- Haslam, N., Bain, P., Douge, L., Lee, M., & Bastian, B. (2005). More human than you: Attributing humanness to self and others. *Journal of Personality and Social Psychology*, 89, 937-950.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Publications.
- Holt-Lunstad, J. (2018). Why social relationships are important for physical health: A systems approach to understanding and modifying risk and protection. *Annual Review of Psychology*, 69, 437-458.
- Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on Aging*, 26, 655-672.
- John, L. K., Barasz, K., & Norton, M. I. (2016). Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences*, 113, 954-959.
- Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, 34, 107-117.
- Jones, E. E., & Archer, R. L. (1976). Are there special effects of personalistic self-disclosure? *Journal of Experimental Social Psychology*, 12, 180-193.
- Kardas, M., Kumar, A., & Epley, N. (2020). *Let it go: Exaggerating the reputational costs of*

revealing secrets creates a misplaced barrier to transparency in relationships.

Manuscript in preparation.

Kashdan, T. B., McKnight, P. E., Fincham, F. D., & Rose, P. (2011). When curiosity breeds intimacy: Taking advantage of intimacy opportunities and transforming boring conversations. *Journal of Personality, 79*, 1369-1402.

Kashdan, T. B., & Roberts, J. E. (2006). Affective outcomes in superficial and intimate interactions: Roles of social anxiety and curiosity. *Journal of Research in Personality, 40*, 140-167.

Keelan, J. P. R., Dion, K. K., & Dion, K. L. (1998). Attachment style and relationship satisfaction: Test of a self-disclosure explanation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 30*, 24-35.

Kelly, A. E., & McKillop, K. J. (1996). Consequences of revealing personal secrets. *Psychological Bulletin, 120*, 450-465.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. The Guilford Press.

Kluger, A. N., Malloy, T. E., Pery, S., Itzhakov, G., Castro, D. R., Lipetz, L., Sela, Y., Turjeman-Levi, Y., Lehmann, M., New, M., & Borut, L. (2020). Dyadic listening in teams: Social relations model. *Applied Psychology: An International Review*.

Kluger, A. N., & Zaidel, K. (2013). Are listeners perceived as leaders? *International Journal of Listening, 27*, 73-84.

Kumar, A., & Epley, N. (2018). Undervaluing gratitude: Expressers misunderstand the consequences of showing appreciation. *Psychological Science, 29*, 1423-1435.

Kumar, A., & Epley, N. (2021). *A little good goes an unexpectedly long way: Underestimating the positive impact of kindness on recipients*. Manuscript in preparation.

- Kumar, A., & Epley, N. (2021). It's surprisingly nice to hear you: Misunderstanding the impact of communication media can lead to suboptimal choices of how to connect with others. *Journal of Experimental Psychology: General*, *150*, 595-607.
- Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*, 371-375.
- Leunissen, J. M., De Cremer, D., van Dijke, M., & Folmer, C. P. R. (2014). Forecasting errors in the averseness of apologizing. *Social Justice Research*, *27*, 322-339.
- Levine, E. E., & Cohen, T. R. (2018). You can handle the truth: Mispredicting the consequences of honest communication. *Journal of Experimental Psychology: General*, *147*(9), 1400-1429.
- Mallett, R. K., Wilson, T. D., & Gilbert, D. T. (2008). Expect the unexpected: Failure to anticipate similarities leads to an intergroup forecasting error. *Journal of Personality and Social Psychology*, *94*, 265-277.
- Maner, J. K., DeWall, C. N., Baumeister, R. F., & Schaller, M. (2007). Does social exclusion motivate interpersonal reconnection? Resolving the "porcupine problem". *Journal of Personality and Social Psychology*, *92*, 42-55.
- Margolis, S., & Lyubomirsky, S. (2020). Experimental manipulation of extraverted and introverted behavior and its effects on well-being. *Journal of Experimental Psychology: General*, *149*(4), 719-731.
- Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations. *Psychological Science*, *21*, 539-541.
- Milek, A., Butler, E. A., Tackman, A. M., Kaplan, D. M., Raison, C. L., Sbarra, D. A., ... &

- Mehl, M. R. (2018). "Eavesdropping on happiness" revisited: A pooled, multisample replication of the association between life satisfaction and observed daily conversation quantity and quality. *Psychological Science, 29*, 1451-1462.
- Miller, N. E. (1944). Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders* (pp. 431-465). New York: Ronald.
- Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*, 6-27.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.
- Omarzu, J. (2000). A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review, 4*, 174-185.
- Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science, 8*, 162-166.
- Pietrzak, J., Downey, G., & Ayduk, O. (2005). Rejection sensitivity as an interpersonal vulnerability. In M. Baldwin (Ed.), *Interpersonal cognition* (pp. 62-84). New York: Guilford.
- Ruan, Y., Reis, H. T., Clark, M. S., Hirsch, J. L., & Bink, B. D. (2020). Can I tell you how I feel? Perceived partner responsiveness encourages emotional expression. *Emotion, 20*(3), 329-342.
- Rubin, Z. (1975). Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology, 11*, 233-260.
- Russell, A. M. T., & Fiske, S. T. (2008). It's all relative: Competition and status drive interpersonal perception. *European Journal of Social Psychology, 38*, 1193-1201.

- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68-78.
- Sandstrom, G. M., & Boothby, E. J. (2021). Why do people avoid talking to strangers? A mini meta-analysis of predicted fears and actual experiences talking to a stranger. *Self & Identity, 20*, 47-71.
- Sandstrom, G. M., & Dunn, E. W. (2014a). Is efficiency overrated? Minimal social interactions lead to belonging and positive affect. *Social Psychological and Personality Science, 5*, 437-442.
- Sandstrom, G. M., & Dunn, E. W. (2014b). Social interactions and well-being: The surprising power of weak ties. *Personality and Social Psychology Bulletin, 40*, 910-922.
- Savitsky, K., Epley, N., & Gilovich, T. (2001). Do others judge us as harshly as we think? Overestimating the impact of our failures, shortcomings, and mishaps. *Journal of Personality and Social Psychology, 81*, 44-56.
- Schopenhauer, A. (1964). (T. B. Saunders, Trans.). *The Pessimist's Handbook*. Lincoln: University of Nebraska Press. (Original work published 1851)
- Schroeder, J., Lyons, D., & Epley, N. (2021). Hello, stranger?: Pleasant conversations are preceded by fears of starting one. *Manuscript under review*.
- Shearer, C. (2017). *More than just talking: The role of self-disclosure in the fast friends procedure*. (Unpublished honors thesis). Seton Hall University, South Orange, USA.
- Slepian, M. L., & Moulton-Tetlock, E. (2019). Confiding secrets and well-being. *Social Psychological and Personality Science, 10*, 472-484.
- 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.
- Strachman, A., & Gable, S. L. (2006). What you want (and do not want) affects what you see (and do not see): Avoidance social goals and social events. *Personality and Social Psychology Bulletin*, 32, 1446-1458.
- Sun, J., Harris, K., & Vazire, S. (2019). Is well-being associated with the quantity and quality of social interactions? *Journal of Personality and Social Psychology*. Advance online publication.
- Taylor, D. A. (1968). The development of interpersonal relationships: Social penetration processes. *The Journal of Social Psychology*, 75, 79-90.
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, 1124, 208-224.
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, 27, 813-833.
- von Hippel, W. (2018). *The Social Leap*. New York, NY: Harper Collins.
- Waytz, A., Schroeder, J., & Epley, N. (2014). The lesser minds problem. In P. G. Bain, J. Vaes, & J.-P. Leyens (Eds.), *Humanness and dehumanization* (pp. 49–67). New York, NY: Psychology Press.
- Wesselmann, E. D., Cardoso, F. D., Slater, S., & Williams, K. D. (2012). To be looked at as though air: Civil attention matters. *Psychological Science*, 23, 166-168.
- Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 78, 821-836.

- Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, *67*, 222-232.
- Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*, 1251-1263.
- Zelenski, J. M., Whelan, D. C., Nealis, L. J., Besner, C. M., Santoro, M. S., & Wynn, J. E. (2013). Personality and affective forecasting: Trait introverts underpredict the hedonic benefits of acting extraverted. *Journal of Personality and Social Psychology*, *104*, 1092-1108.
- Zhao, X. & Epley, N. (in press). Insufficiently complimentary?: Underestimating the positive impact of compliments creates a barrier to expressing them. *Journal of Personality and Social Psychology*.

Appendix

Intimacy Ratings by Question

Index	Question	M_{intimacy} (SD_{intimacy})
S01	What do you think about the weather today?	1.21 (2.11)
S02	How often do you come here?	1.97 (2.06)
S03	How did you celebrate last Halloween?	2.14 (2.09)
S04	How often do you get your hair cut? Where do you go? Have you ever had a really bad haircut experience?	2.39 (2.13)
S05	What is the best TV show you've seen in the last month? Tell your partner about it.	2.68 (2.31)
S06	When was the last time you walked for more than an hour? Describe where you went and what you saw.	2.69 (2.19)
S07	Do you like to get up early or stay up late? Why?	2.89 (1.93)
S08	Do you have anything planned for later today? When are you going to do it?	3.22 (2.01)
S09	Can you describe a conversation you had with another person earlier today?	3.30 (2.14)
S10	What's your daily routine like?	4.40 (1.85)
D01	What would constitute a "perfect" day for you?	4.75 (2.38)
D02	Where is somewhere you've visited that you felt really had an impact on who you are today?	5.24 (1.94)
D03	If you were going to become a close friend with the other participant, please share what would be important for him or her to know.	6.35 (1.99)
D04	If a crystal ball could tell you the truth about yourself, your life, the future, or anything else, what would you want to know?	6.38 (2.12)
D05	For what in your life do you feel most grateful? Tell the other participant about it.	6.47 (1.96)
D06	Is there something you've dreamed of doing for a long time? Why haven't you done it?	6.50 (2.10)
D07	What is one of the more embarrassing moments in your life?	6.83 (1.88)
D08	What is one of your most meaningful memories? Why is it meaningful for you?	7.29 (2.02)
D09	Can you describe a time you cried in front of another person?	7.39 (2.16)
D10	If you could undo one mistake you have made in your life, what would it be and why would you undo it?	7.82 (1.98)

Note. Intimacy ratings by question. Numbers outside parentheses denote means; numbers inside parentheses denote standard deviations. S01 through S10 denote shallow questions in ascending order of intimacy. D01 through D10 denote deep questions in ascending order of intimacy.

Supporting Information For:

Overly Shallow?:

Miscalibrated Expectations Create a Barrier to Deeper Conversation

Pre-Test for Conversation Questions

Method

Participants. Participants were recruited from Amazon's Mechanical Turk ($N = 45$; $M_{\text{age}} = 35.71$; $SD_{\text{age}} = 9.58$; 31.11% female; 82.22% Caucasian) to complete the study in exchange for \$2.00. An additional 3 participants were excluded because they failed one or more attention checks.

Procedure. Participants imagined visiting a “social interaction” research lab where they would engage in a get-to-know-you discussion with another study participant who they had never met before. They were told that this discussion would be called the “sharing game.”

Then participants viewed each of the 20 discussion questions in randomized order (see Table S1). They responded to four items for each discussion question: “How much do you think your answers would reveal about your identity—about who you are as a person?” (0 = *nothing at all*; 10 = *quite a bit*); “Questions can range from superficial and shallow to intimate and deep. How intimate do you consider these questions to be?” (0 = *not intimate at all*; 10 = *extremely intimate*); “To what extent do you think these questions would require you to reveal meaningful information about yourself?” (0 = *not at all*; 10 = *quite a bit*); and “How vulnerable do you think you would feel while asking and answering these questions with a stranger?” (0 = *not at all vulnerable*; 10 = *extremely vulnerable*).

Then participants completed two attention checks: they reported who they were asked to imagine speaking with (*A stranger who I'd never met before* vs. *A friend with whom I'd lost touch*) and they reported whether they paid attention throughout the study (*Yes* vs. *No*). Finally, participants reported demographic information and received payment.

Results

The four items were highly correlated ($\alpha = .96$) and so we collapsed across them to form an intimacy scale by averaging participants' responses (see Table S1).

Index	Question	<i>M (SD)</i>
S01	What do you think about the weather today?	1.21 (2.11)
S02	How often do you come here?	1.97 (2.06)
S03	How did you celebrate last Halloween?	2.14 (2.09)
S04	How often do you get your hair cut? Where do you go? Have you ever had a really bad haircut experience?	2.39 (2.13)
S05	What is the best TV show you've seen in the last month? Tell your partner about it.	2.68 (2.31)
S06	When was the last time you walked for more than an hour? Describe where you went and what you saw.	2.69 (2.19)
S07	Do you like to get up early or stay up late? Why?	2.89 (1.93)
S08	Do you have anything planned for later today? When are you going to do it?	3.22 (2.01)
S09	Can you describe a conversation you had with another person earlier today?	3.30 (2.14)
S10	What's your daily routine like?	4.40 (1.85)
D01	What would constitute a "perfect" day for you?	4.75 (2.38)
D02	Where is somewhere you've visited that you felt really had an impact on who you are today?	5.24 (1.94)
D03	If you were going to become a close friend with the other participant, please share what would be important for him or her to know.	6.35 (1.99)
D04	If a crystal ball could tell you the truth about yourself, your life, the future, or anything else, what would you want to know?	6.38 (2.12)
D05	For what in your life do you feel most grateful? Tell the other participant about it.	6.47 (1.96)
D06	Is there something you've dreamed of doing for a long time? Why haven't you done it?	6.50 (2.10)
D07	What is one of the more embarrassing moments in your life?	6.83 (1.88)
D08	What is one of your most meaningful memories? Why is it meaningful for you?	7.29 (2.02)
D09	Can you describe a time you cried in front of another person?	7.39 (2.16)
D10	If you could undo one mistake you have made in your life, what would it be and why would you undo it?	7.82 (1.98)

Table S1. Intimacy ratings by question. Numbers outside parentheses denote means; numbers inside parentheses denote standard deviations. S01 through S10 denote shallow questions in ascending order of intimacy. D01 through D10 denote deep questions in ascending order of intimacy.

Introduction #1 (Listener vs. Asker vs. Discussant vs. Answerer)

Method

Participants. Participants were recruited from Amazon's Mechanical Turk ($N = 188$ individuals after exclusions; $M_{\text{age}} = 34.91$; $SD_{\text{age}} = 10.66$; 46.28% female; 66.49% Caucasian) to complete the study in exchange for \$0.50. An additional 13 participants were excluded for failing the attention check.

Procedure. Participants imagined visiting a “social interaction” research lab where they would speak with another study participant who they had never met before. They were told that this discussion would be called the “sharing game.”

Then participants were randomly assigned to one of four roles: Listener, Asker, Discussant, or Answerer. Listeners were told that “the other study participant would draw 5 questions out of a hat and answer them, and you would listen to his or her answers.” Askers were told that they would “ask 5 questions and the other participant would answer these 5 questions while you simply listen.” Discussants were told that they and the other study participant “would ask and answer 5 questions while interacting with one another. That is, both you and the other study participant would answer the same 5 questions.” Answerers were told that “the other participant would ask 5 questions and you would answer these 5 questions while the other participant simply listens.”

Then participants viewed the list of 20 pre-tested discussion questions in randomized order. They selected the 5 questions that they would most like to listen to, ask, discuss, or answer during the sharing game. Then participants completed one attention check item in which they reported their role in the sharing game (Listener vs. Asker vs. Discussant vs. Answerer). Finally, participants reported demographic information and received payment.

Results

Based on the pre-test ratings, we computed the mean intimacy of the questions that participants selected in each condition. The mean intimacy of the questions selected varied significantly by condition, $F(3, 184) = 15.02, p < .001, \eta_p^2 = .20$. Consistent with our predictions, Listeners ($M = 5.06, SD = 1.33$) chose more intimate questions than Discussants ($M = 4.04, SD = 1.09$), $t(184) = -4.10, p < .001, 95\% CI_{\text{difference}} = [-1.51, -0.53], d = -0.86$, or Answerers ($M = 3.58, SD = 0.91$), $t(184) = -6.13, p < .001, 95\% CI_{\text{difference}} = [-1.96, -1.00], d = -1.25$. Likewise, Askers ($M = 4.71, SD = 1.38$) chose more intimate questions than Discussants, $t(184) = -2.69, p = .008, 95\% CI_{\text{difference}} = [-1.16, -0.18], d = -0.56$, or Answerers, $t(184) = -4.67, p < .001, 95\% CI_{\text{difference}} = [-1.61, -0.65], d = -0.95$.

The number of deep questions selected also varied by condition, $F(3, 184) = 14.54, p < .001, \eta_p^2 = .19$. Listeners ($M = 3.06$ out of 5, $SD = 1.57$) chose more deep questions than Discussants ($M = 1.89, SD = 1.30$), $t(184) = -3.97, p < .001, 95\% CI_{\text{difference}} = [-0.35, -0.12], d = -0.83$, or Answerers ($M = 1.36, SD = 1.16$), $t(184) = -5.92, p < .001, 95\% CI_{\text{difference}} = [-0.45, -0.23], d = -1.20$. Likewise, Askers ($M = 2.74, SD = 1.61$) chose more deep questions than Discussants, $t(184) = 2.86, p = .005, 95\% CI_{\text{difference}} = [0.05, 0.29], d = 0.60$, or Answerers, $t(184) = 4.76, p < .001, 95\% CI_{\text{difference}} = [0.16, 0.39], d = 0.97$.

Introduction #2 (Listener vs. Answerer vs. Revealer)

Method

Participants. Participants were recruited from Amazon's Mechanical Turk ($N = 144$; $M_{\text{age}} = 36.28$; $SD_{\text{age}} = 10.45$; 43.06% female; 83.33% Caucasian) to complete the study in exchange for \$1.25. An additional 6 participants were excluded for failing the attention check.

Procedure. The procedure was identical to the experiment discussed above with two exceptions. First, participants were randomly assigned to one of three roles: Listener, Answerer, or Revealer. Listeners and Answerers received instructions analogous to those from the previous experiment. Revealers were told that “you would write down your answers to all of the questions and then the other study participant would draw 5 questions out of a hat. Then you would reveal your answers to these 5 questions while the other participant simply listens.” Revealers were asked to imagine that they had already written down their answers to each of the 20 questions before they chose the 5 that they preferred for the sharing game.

Second, after participants selected 5 out of 20 questions, we added exploratory items in which participants evaluated each of the 20 questions individually by reporting how easy it would be for them (or the other person) to come up with a response to the question (0 = *not at all easy*; 10 = *extremely easy*) and how intimate they considered the question to be (0 = *not intimate at all*; 10 = *extremely intimate*).

Results

Based on the pre-test ratings, we computed the mean intimacy of the questions that participants selected in each condition. The mean intimacy of the questions selected varied significantly by condition, $F(2, 141) = 11.89, p < .001, \eta_p^2 = .14$. Answerers ($M = 3.79, SD = 1.06$) and Revealers ($M = 3.83, SD = 1.06$) selected similarly intimate questions, $t(141) = 0.14, p$

= .885, 95% $CI_{\text{difference}} = [-0.43, 0.49]$, $d = 0.03$, but consistent with our predictions, Listeners ($M = 4.80$, $SD = 1.30$) selected significantly more intimate questions than either Answerers, $t(141) = 4.33$, $p < .001$, 95% $CI_{\text{difference}} = [0.55, 1.47]$, $d = 0.88$, or Revealers, $t(141) = -4.13$, $p < .001$, 95% $CI_{\text{difference}} = [-1.44, -0.51]$, $d = -0.85$.

Furthermore, the number of deep questions that participants selected varied significantly by condition, $F(2, 141) = 12.71$, $p < .001$, $\eta_p^2 = .15$. Listeners ($M = 2.85$ out of 5, $SD = 1.55$) chose more deep questions than either Answerers ($M = 1.62$, $SD = 1.23$), $t(141) = -4.52$, $p < .001$, 95% $CI_{\text{difference}} = [-0.35, -0.14]$, $d = -0.92$, or Revealers ($M = 1.68$, $SD = 1.24$), $t(141) = 4.23$, $p < .001$, 95% $CI_{\text{difference}} = [0.12, 0.34]$, $d = 0.87$. The number of deep questions selected did not differ significantly between Answerers and Revealers, $t(141) = -0.22$, $p = .824$, 95% $CI_{\text{difference}} = [-0.12, 0.10]$, $d = -0.05$.

Experiment 1a

Full Output from ANOVA in Main Text

In a 2 (measurement phase: expectations, experiences) \times 2 (target: self, partner) ANOVA with repeated measures on both factors and interest as the dependent measure, we observed a significant main effect of measurement phase, $F(1, 24) = 79.62, p < .001, \eta_p^2 = .77$, such that participants underestimated interest, and a significant main effect of target, $F(1, 24) = 121.60, p < .001, \eta_p^2 = .84$, such that participants' own interest was greater than their partners' interest. These main effects were qualified by a significant measurement phase \times target interaction effect, $F(1, 24) = 14.63, p < .001, \eta_p^2 = .38$, such that participants underestimated their partner's interest more than they underestimated their own.

Mediational Analyses at the Level of the Individual

We conducted mediational analyses at the level of the individual, with measurement phase (expectations vs. experiences) as the independent variable, partner interest as the mediating variable, and the primary measures as dependent variables in separate analyses. The indirect effects were non-significant for awkwardness ($b = -0.55, 95\% \text{ CI} = [-1.83, 0.57]$) and happiness ($b = -0.83, 95\% \text{ CI} = [-1.90, 0.15]$) but significant for connectedness ($b = -0.99, 95\% \text{ CI} = [-1.45, -0.62]$).

Miscalibrated Beliefs vs. Response Bias

Two sets of additional analyses suggest that differences between participants' expectations and experience reflect miscalibrated beliefs and not response biases.

First, we did not find evidence that participants inflated their experience ratings to rationalize or justify a somewhat awkward conversation. Across our dyadic experiments, participants who rated their conversations to be especially awkward tended to rate their

conversation partners as less caring in both the shallower ($b = -0.18, SE = 0.04, t(1193.98) = -4.72, p < .001, 95\% CI = [-0.26, -0.11]$) and deeper conditions ($b = -0.12, SE = 0.03, t(1193.81) = -4.27, p < .001, 95\% CI = [-0.18, -0.07]$), tended to report feeling less connected to their conversation partners in both the shallower ($b = -0.22, SE = 0.03, t(1630.18) = -7.47, p < .001, 95\% CI = [-0.28, -0.17]$) and deeper conditions ($b = -0.17, SE = 0.02, t(1630.54) = -7.35, p < .001, 95\% CI = [-0.22, -0.13]$), and tended to report less happiness in both the shallower ($b = -0.23, SE = 0.03, t(1390.52) = -7.63, p < .001, 95\% CI = [-0.29, -0.17]$) and deeper conditions ($b = -0.20, SE = 0.02, t(1391.43) = -8.45, p < .001, 95\% CI = [-0.25, -0.16]$).

Second, across our dyadic experiments, paired participants' experience ratings tended to be positively correlated for each of the primary measures: partner care or interest ($b = 0.11, SE = 0.03, t(995.57) = 3.65, p < .001, 95\% CI = [0.05, 0.18]$), awkwardness ($b = 0.07, SE = 0.03, t(1335.98) = 2.44, p = .015$), connectedness or enjoyment ($b = 0.21, SE = 0.03, t(1334.22) = 7.57, p < .001$), and happiness ($b = 0.08, SE = 0.03, t(995.89) = 2.41, p = .016$).

Gender Effects

The amount of miscalibration did not differ significantly between same-gender and mixed-gender pairs on the own-interest, partner-interest, awkwardness, or connectedness measures, $F_s(1, 23) \leq 0.90, p_s \geq .352, \eta_p^2_s \leq .04$. Mixed-gender pairs were significantly more likely than same-gender pairs to underestimate their happiness, $F(1, 23) = 4.75, p = .040, \eta_p^2 = .17$.

Experiment 1b

Method

Participants. Fifty pairs of employees and executives at a financial services firm were recruited during two presentations delivered on the same day at the firm ($N = 100$ individuals; $M_{\text{age}} = 39.57$; $SD_{\text{age}} = 10.57$; 63.00% female; 81.00% Caucasian). No employees participated in both sessions and all employees entered the sessions unaware that they would be participating in any experiment. An additional 11 participants could not be matched with their partner after the session because either 1 or 3 participants reported the same pair ID, and so were removed from analyses. Another 6 individuals listened to the presentation remotely and did not speak with another person; their survey responses also were not analyzed.

Procedure. The procedure was nearly identical to Experiment 1a with several minor changes. First, we modified the wording of the fourth discussion question from, “Can you describe a time you cried in front of another person?” to “Can you tell me about one of the last times you cried in front of another person?” Second, rather than report how interested they would be in their partner’s responses (and vice versa), participants instead reported how much they would care about their partner’s responses (0 = *not at all*; 10 = *quite a bit*) and how much their partner would care about the participant’s responses (0 = *not at all*; 10 = *quite a bit*). Third, after reporting their own experiences on the primary measures, participants indicated the degree to which they responded to all the questions openly and honestly (0 = *not at all*; 10 = *completely*), and completed an analogous measure for their partner after reporting their perceptions of their partner’s experiences. Fourth, participants completed a free-response item at the end of the survey in which they reported how the conversation made them feel. These free-responses were used only for illustrative purposes and are not analyzed quantitatively. Finally, whereas

participants in Experiment 1a separately reported how often they wished they would engage in small talk and how often they wished they would engage in deep conversation, participants in this experiment reported how deep they wished their conversations in the office would be (-5 = *much less deep than they usually are*; 0 = *neither more nor less deep than they usually are*; 5 = *much deeper than they usually are*).

Results

The strength-of-bond and liking items were highly correlated in both expectations ($\alpha = .81$) and experiences ($\alpha = .72$) and so we again combined these items to form a connectedness scale.

The employees who participated in the two sessions held different positions in the firm—those in the morning session worked at the management or executive level, whereas those in the afternoon session included a broader mix of mid-level managers and non-management employees—and so we first tested whether the findings from the two sessions differed in any meaningful way. To do this, we performed a 2 (measurement phase: expectations, experiences) \times 2 (session: morning, afternoon) ANOVA for each of the primary measures separately. The measurement phase \times session interaction effects were nonsignificant on all items, $F_s(1, 48) \leq 0.18$, $p_s \geq .670$, $\eta_p^2_s \leq .004$, indicating that the amount of miscalibration on the primary measures did not vary significantly by session. We therefore analyzed both sessions together as one experiment.

Participants underestimated both how much they would care about their partner's responses and how much their partner would care about the participant's responses, but consistent with our hypotheses, participants especially underestimated their partner's care. These patterns were confirmed in a 2 (measurement phase: expectations, experiences) \times 2 (target: self,

partner) ANOVA with repeated measures on both factors. We observed a significant main effect of measurement phase, $F(1, 49) = 154.62, p < .001, \eta_p^2 = .76$, such that participants underestimated care, a significant main effect of target, $F(1, 49) = 40.31, p < .001, \eta_p^2 = .45$, such that participants' own care was greater than their partners' care, and a significant measurement phase \times target interaction effect, $F(1, 49) = 16.26, p < .001, \eta_p^2 = .25$.

To decompose this interaction effect, we tested for differences between predicted and actual care for each measure separately. Participants tended to underestimate how much they would care about their partner's responses, *paired* $t(49) = -10.91, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.97, -1.35], d = -1.40$, and how much they would perceive their partner to care about one's own responses, *paired* $t(49) = -11.74, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.62, -1.86], d = -1.72$.

Further, our data suggest that these differences between expectations and experiences reflect miscalibrated beliefs about a deep conversation and not response biases that might arise in the ratings of their partner's care after the conversation. Participants' expectations of their partner's care underestimated both their partner's expected care before the conversation, *paired* $t(49) = -6.26, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.23, -0.63], d = -0.72$, as well as their partner's self-reported care after the conversation, *paired* $t(49) = -14.62, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.95, -2.23], d = -2.07$.

Participants' conversations also unfolded significantly better than they anticipated: Participants overestimated awkwardness, *paired* $t(49) = 12.11, p < .001, 95\% \text{ CI}_{\text{difference}} = [2.34, 3.26], d = 1.71$, underestimated connectedness, *paired* $t(49) = -10.82, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.05, -1.41], d = -1.53$, and underestimated happiness, *paired* $t(49) = -11.50, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.27, -1.59], d = -1.63$.

We hypothesized that underestimating others' care would explain miscalibration on the primary measures. In within-pairs mediational analyses, the indirect effects were non-significant for awkwardness ($b = 0.02$, 95% CI = [-0.66, 0.82]) but significant for both connectedness ($b = -0.67$, 95% CI = [-1.24, -0.18]) and happiness ($b = -0.85$, 95% CI = [-1.43, -0.35]). Thus, we observed significant mediation for connectedness in both Experiments 1a and 1b, but for awkwardness only in Experiment 1a and for happiness only in Experiment 1b. We continue testing the mediating role of partner care in the following experiments.

Participants wished that they engaged in relatively deeper conversations in the office than they did typically ($M = 2.08$, $SD = 1.72$), $t(99) = 12.12$, $p < .001$, 95% CI = [1.74, 2.42], $d = 1.21$. Furthermore, participants reported being highly open and honest during the conversations ($M = 9.20$, $SD = 1.02$) and perceived their partners to have been highly open and honest as well ($M = 8.92$, $SD = 1.03$), suggesting that participants engaged in relatively intimate conversations as intended.

Mediational Analyses at the Level of the Individual

We conducted mediational analyses at the level of the individual, with measurement phase (expectations vs. experiences) as the independent variable, partner care as the mediating variable, and the primary measures as dependent variables in separate analyses. The indirect effects were non-significant for awkwardness ($b = 0.31$, 95% CI = [-0.27, 0.91]) but significant for both connectedness ($b = -0.77$, 95% CI = [-1.21, -0.34]) and happiness ($b = -1.04$, 95% CI = [-1.55, -0.62]).

Gender Effects

The amount of miscalibration did not differ significantly between same-gender and mixed-gender pairs on any measure, $F_s(1, 48) \leq 1.57$, $p_s \geq .217$, $\eta_p^2_s \leq .03$.

Experiment 1c

Method

Participants. Fifty-six groups of two to three individuals were recruited during a master's degree student orientation event split into two sessions ($N = 137$ individuals after exclusions; $M_{\text{age}} = 36.95$; $SD_{\text{age}} = 4.91$; 25.55% female; 35.04% Caucasian). No students participated in both sessions and all students entered the sessions unaware that they would be participating in any experiment. An additional 35 participants were excluded because one or more participants in the group reported technical difficulties during the conversation. Further, we removed another 11 groups from analyses because one or more group members did not consent to have their data analyzed for scientific research.

Procedure. The procedure was nearly identical to Experiment 1a with several notable exceptions. First, we conducted the research study (and the rest of the orientation event) virtually through the Zoom video conferencing software due to the COVID-19 pandemic. Participants engaged in their conversations in digital “breakout rooms” in which only their conversation partners could be seen and heard. Although incidental to our primary hypotheses, the video conferencing format allows us to test the generalizability of our findings by having participants interact across a different communication medium. Second, we created groups of two to three participants—rather than strictly creating dyads—due to the limited number of breakout rooms available through the video conferencing software. Third, we removed the own interest and partner interest measures but retained the awkwardness, strength of bond, and liking measures. Fourth, we measured predicted and actual enjoyment rather than happiness. Fifth, we slightly modified the deep conversation topics. Participants received the following questions:

1. If I were to become a good friend of yours, what would be most important for me to know?
2. For what in your life do you feel most grateful? Please tell me about it.
3. Can you tell me about one of the last times you cried in front of another person?
4. What are you most anxious about, right now, starting this program at [removed for blinded review]?

Sixth, participants reported their own experiences after the conversation but did not report their perceptions of their partner's experiences—with the exception that they reported both their own openness and honesty and their perception of their partners' openness and honesty after the other experience measures. Seventh, after the experience measures, participants reported the intimacy of their conversation, the intimacy of their typical conversations with strangers, and the intimacy of the conversations they wish they would have with strangers, on separate scales from 0 (*light conversation and small talk*) to 10 (*meaningful conversation and deep talk*). Finally, because participants conversed through a video call, they indicated at the end of the survey whether they experienced any technological difficulties during the conversation.

Results

We first tested whether the findings from the two sessions differed in any meaningful way. To do this, we performed a 2 (measurement phase: expectations, experiences) \times 2 (session: morning, afternoon) ANOVA for the awkwardness and connectedness measures separately. The measurement phase \times session interaction effects were nonsignificant on both items, $F_s(1, 54) \leq 0.54$, $p_s \geq .467$, $\eta_p^2_s \leq .01$, indicating that the magnitude of miscalibration on the primary measures did not vary significantly by session. We therefore collapsed across both sessions in

the following analyses. The strength-of-bond, liking, and enjoyment items were highly correlated ($\alpha_{\text{predictions}} = .71$, $\alpha_{\text{experiences}} = .77$), and so we collapsed them to form a connectedness scale.

Replicating the findings of the Experiments 1a-b, participants' conversations unfolded substantially better than they anticipated: Participants overestimated how awkward they would feel during their conversations, *paired* $t(55) = 9.05$, $p < .001$, 95% $CI_{\text{difference}} = [1.78, 2.79]$, $d = 1.39$, but underestimated how connected they would feel to their conversation partner, *paired* $t(55) = -10.13$, $p < .001$, 95% $CI_{\text{difference}} = [-1.92, -1.29]$, $d = -1.69$.

Participants reported that their conversations were relatively deep and meaningful ($M = 6.84$, $SD = 1.40$), and that both they ($M = 9.09$, $SD = 0.85$) and their conversation partners ($M = 8.76$, $SD = 0.85$) were highly open and honest during the conversation. Participants also indicated that their typical conversations with strangers ($M = 3.60$, $SD = 2.31$) were significantly less intimate than they wished they would be ($M = 5.81$, $SD = 2.37$), *paired* $t(136) = -11.32$, $p < .001$, 95% $CI_{\text{difference}} = [-2.60, -1.83]$, $d = -0.94$.

Analyses Without Exclusions

We reanalyzed the data with all 172 participants (71 groups) included. Consistent with the main analyses already described, participants overestimated how awkward they would feel during their conversations, *paired* $t(70) = 10.69$, $p < .001$, 95% $CI_{\text{difference}} = [1.85, 2.69]$, $d = 1.42$, but underestimated how connected they would feel to their conversation partner, *paired* $t(70) = -11.66$, $p < .001$, 95% $CI_{\text{difference}} = [-1.82, -1.29]$, $d = -1.69$.

Participants reported that their conversations were relatively deep and meaningful ($M = 6.81$, $SD = 1.48$), and that both they ($M = 9.02$, $SD = 0.88$) and their conversation partner ($M = 8.75$, $SD = 0.91$) were highly open and honest during the conversation. Participants also indicated that their typical conversations with strangers ($M = 3.63$, $SD = 2.39$) were significantly

less deep and meaningful than they wished they would be ($M = 5.87$, $SD = 2.40$), *paired* $t(171) = -12.06$, $p < .001$, 95% $CI_{\text{difference}} = [-2.60, -1.87]$, $d = -0.93$.

Gender Effects

The amount of miscalibration did not differ significantly between single-gender and mixed-gender groups on either awkwardness or connectedness, $F_s(1, 54) \leq 3.19$, $p_s \geq .080$, η_p^2 s $\leq .06$.

Experiment 2

Full Output from ANOVAs in Main Text

For awkwardness, a 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA on evaluations of awkwardness with repeated measures on the second factor yielded a significant main effect of conversation, $F(1, 87) = 100.25, p < .001, \eta_p^2 = .54$, indicating greater awkwardness in the deep condition, and a significant main effect of measurement phase, $F(1, 87) = 178.31, p < .001, \eta_p^2 = .67$, indicating that participants overestimated awkwardness. These main effects were qualified by the predicted interaction, $F(1, 87) = 71.34, p < .001, \eta_p^2 = .45$, indicating that participants in the deep condition overestimated awkwardness more than participants in the shallow condition.

For connectedness, a 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with the connectedness scale as the dependent measure yielded a significant main effect of conversation, $F(1, 87) = 4.76, p = .032, \eta_p^2 = .05$, such that participants in the deep condition reported greater connectedness than participants in the shallow condition, and a significant main effect of measurement phase, $F(1, 87) = 98.24, p < .001, \eta_p^2 = .53$, such that participants underestimated connectedness across conditions. The conversation \times measurement phase interaction effect was non-significant, $F(1, 87) = 0.12, p = .734, \eta_p^2 = .001$.

Moderation by Big Five Personality Traits

We tested whether personality moderated our findings by regressing overestimation of awkwardness, and underestimation of connectedness, simultaneously over the Big Five traits. In the shallow condition, overestimation of awkwardness was not significantly associated with agreeableness ($\beta = -0.01, t(98) = -0.07, p = .947$), conscientiousness ($\beta = -0.09, t(98) = -0.88, p = .379$), extraversion ($\beta = -0.03, t(98) = -0.34, p = .738$), or openness ($\beta = 0.11, t(98) = 1.06, p =$

.290), but was significantly associated with higher neuroticism ($\beta = -0.26$, $t(98) = -2.58$, $p = .011$). Underestimation of connectedness was not significantly associated with agreeableness ($\beta = 0.04$, $t(98) = 0.43$, $p = .669$), conscientiousness ($\beta = -0.09$, $t(98) = -0.90$, $p = .372$), or extraversion ($\beta = 0.03$, $t(98) = 0.27$, $p = .791$), but was significantly associated with higher neuroticism ($\beta = 0.22$, $t(98) = 2.19$, $p = .031$) and lower openness ($\beta = -0.21$, $t(98) = -2.16$, $p = .033$).

In the deep condition, overestimation of awkwardness was not significantly associated with any Big Five personality traits: agreeableness ($\beta = -0.17$, $t(68) = -1.36$, $p = .179$), conscientiousness ($\beta = 0.06$, $t(68) = 0.51$, $p = .610$), extraversion ($\beta = 0.16$, $t(68) = 1.28$, $p = .204$), neuroticism ($\beta = -0.02$, $t(68) = -0.20$, $p = .845$), or openness ($\beta = -0.04$, $t(68) = -0.31$, $p = .757$). Underestimation of connectedness likewise was not significantly associated with any Big Five personality traits: agreeableness ($\beta = 0.10$, $t(68) = 0.83$, $p = .409$), conscientiousness ($\beta = -0.18$, $t(68) = -1.44$, $p = .153$), extraversion ($\beta = -0.12$, $t(68) = -0.98$, $p = .330$), neuroticism ($\beta = -0.07$, $t(68) = -0.59$, $p = .557$), or openness ($\beta = -0.07$, $t(68) = -0.58$, $p = .561$).

Gender Effects

Neither the amount of miscalibration on the primary measures, nor differences in miscalibration between the shallow and deep conditions, differed significantly between same-gender and mixed-gender pairs, $F_s(1, 85) \leq 2.95$, $p_s \geq .090$, $\eta_p^2_s \leq .03$.

Experiment 3 (Manipulation Check)

To test whether the intimacy manipulation was effective, we recruited a separate group of participants to rate the intimacy of the participant-generated discussion questions.

Method

Participants. Participants were recruited from Amazon's Mechanical Turk ($N = 409$; $M_{\text{age}} = 34.69$; $SD_{\text{age}} = 11.85$; 40.34% female; 68.95% Caucasian) to complete the study in exchange for \$0.50. No participants were excluded. We targeted 400 participants so that each participant-generated discussion question would be evaluated by approximately two raters.

Procedure. Participants imagined visiting a research lab to participate in a "sharing game" with another study participant who they hadn't met before. They were told that they would read a set of questions that they would ask and answer during the sharing game and that they should imagine how they would respond to each question after reading.

Then participants viewed a set of 5 control and 5 deep questions. For each question, participants reported how intimate the question was ($0 = \textit{not intimate at all}$; $10 = \textit{extremely intimate}$). Each participant evaluated the 10 questions that were generated by one participant in the lab experiment, regardless of whether this lab participant was a Writer (whose questions were ultimately used in the discussion) or a Receiver (whose questions were ultimately not used in the discussion). After rating the questions, all participants reported demographic information and received payment.

Results

We computed mean intimacy ratings across the five control questions and five deep questions, separately, for each participant in the original laboratory experiment. Participants discussed more intimate questions in the deep condition ($M = 6.28$, $SD = 1.50$) than in the

control condition ($M = 4.74$, $SD = 1.67$), $t(96) = -4.81$, $p < .001$, 95% $CI_{\text{difference}} = [-2.17, -0.90]$, $d = -0.97$, suggesting that the intimacy manipulation was effective.

Next, we performed the same analysis for all questions that participants generated in Experiment 3. The deep questions ($M = 6.46$, $SD = 1.43$) were rated as more intimate than control questions ($M = 4.71$, $SD = 1.74$), $t(395) = -10.99$, $p < .001$, 95% $CI_{\text{difference}} = [-2.07, -1.44]$, $d = -1.10$, again suggesting that the intimacy manipulation was effective.

Experiment 3

Full Output from ANOVAs in Main Text

A 2 (conversation: control, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the second factor and awkwardness as the dependent measure yielded a marginally significant main effect of conversation, $F(1, 98) = 2.89, p = .092, \eta_p^2 = .03$, such that participants in the deep condition reported directionally greater awkwardness, and a significant main effect of measurement phase, $F(1, 98) = 58.60, p < .001, \eta_p^2 = .37$, such that participants overestimated awkwardness. We also observed a marginally significant conversation \times measurement phase interaction effect, $F(1, 98) = 3.74, p = .056, \eta_p^2 = .04$, such that the participants in the deep condition overestimated awkwardness more than participants in the control condition.

For connectedness, we found a main effect of conversation, $F(1, 98) = 8.67, p = .004, \eta_p^2 = .08$, such that participants reported greater connectedness in the deep condition than in the control condition, and also a main effect of measurement phase, $F(1, 98) = 40.46, p < .001, \eta_p^2 = .29$, such that participants underestimated how connected they would feel to their partner after the conversation. The conversation \times measurement phase interaction effect was non-significant, $F(1, 98) = 0.01, p = .936, \eta_p^2 = .0001$.

For happiness, we found a non-significant main effect of conversation, $F(1, 98) = 2.06, p = .154, \eta_p^2 = .02$, and a significant main effect of measurement phase, $F(1, 98) = 45.07, p < .001, \eta_p^2 = .32$, such that participants underestimated happiness. The conversation \times measurement phase interaction was non-significant, $F(1, 98) = 1.84, p = .179, \eta_p^2 = .02$.

Moderation by Big Five Personality Traits

We tested whether personality moderated our findings by regressing overestimation of awkwardness, and underestimation of connectedness and happiness, simultaneously over the Big Five traits. Among participants in the control condition, overestimation of awkwardness was not significantly associated with any Big Five traits: agreeableness ($\beta = -0.01$, $t(94) = -0.07$, $p = .944$), conscientiousness ($\beta = -0.02$, $t(94) = -0.02$, $p = .985$), extraversion ($\beta = -0.08$, $t(94) = -0.71$, $p = .479$), neuroticism ($\beta = -0.13$, $t(94) = -1.05$, $p = .296$), or openness ($\beta = 0.02$, $t(94) = -0.15$, $p = .883$). Among participants in the deep condition, overestimation of awkwardness was not significantly associated with agreeableness ($\beta = -0.06$, $t(94) = -0.62$, $p = .539$), conscientiousness ($\beta = 0.01$, $t(94) = 0.14$, $p = .891$), extraversion ($\beta = -0.02$, $t(94) = -0.18$, $p = .857$), or neuroticism ($\beta = -0.004$, $t(94) = -0.04$, $p = .970$), but was significantly associated with lower openness ($\beta = 0.45$, $t(94) = 4.34$, $p < .001$).

Among participants in the control condition, underestimation of connectedness was not significantly associated with any of the Big Five traits: agreeableness ($\beta = 0.07$, $t(94) = 0.61$, $p = .544$), conscientiousness ($\beta = 0.21$, $t(94) = 1.77$, $p = .080$), extraversion ($\beta = 0.008$, $t(94) = 0.08$, $p = .940$), neuroticism ($\beta = 0.19$, $t(94) = 1.50$, $p = .137$), or openness ($\beta = 0.002$, $t(94) = 0.02$, $p = .986$). Likewise, among participants in the deep condition, underestimation of connectedness was not significantly associated with any of the Big Five traits: agreeableness ($\beta = -0.03$, $t(94) = -0.26$, $p = .796$), conscientiousness ($\beta = -0.03$, $t(94) = -0.17$, $p = .866$), extraversion ($\beta = 0.18$, $t(94) = 1.64$, $p = .104$), neuroticism ($\beta = -0.18$, $t(94) = -1.68$, $p = .095$), or openness ($\beta = -0.20$, $t(94) = -1.79$, $p = .077$).

Among participants in the control condition, underestimation of happiness was not significantly associated with agreeableness ($\beta = 0.06$, $t(94) = 0.50$, $p = .618$), extraversion ($\beta = 0.12$, $t(94) = 1.07$, $p = .289$), neuroticism ($\beta = 0.20$, $t(94) = 1.59$, $p = .115$), or openness ($\beta = -$

0.06, $t(94) = -0.51, p = .614$), but was significantly associated with higher conscientiousness ($\beta = 0.23, t(94) = 2.01, p = .048$). Among participants in the deep condition, underestimation of happiness was not significantly associated with any Big Five traits: agreeableness ($\beta = 0.04, t(94) = 0.35, p = .729$), conscientiousness ($\beta = 0.08, t(94) = 0.72, p = .474$), extraversion ($\beta = -0.06, t(94) = 0.55, p = .586$), neuroticism ($\beta = -0.10, t(94) = -0.89, p = .378$), or openness ($\beta = -0.14, t(94) = -1.22, p = .226$).

Analyses Without Exclusions

Next, we performed the primary analyses with all 103 pairs of participants included. We performed 2 (conversation: control, deep) \times 2 (measurement phase: expectations, experiences) ANOVAs, with repeated measures on the second factor, for awkwardness, connectedness, and happiness.

Awkwardness. We observed a significant main effect of conversation, $F(1, 101) = 4.44, p = .038, \eta_p^2 = .04$, such that predicted or actual awkwardness was greater for deep questions than control questions, a main effect of measurement phase, $F(1, 101) = 57.44, p < .001, \eta_p^2 = .36$, such that participants overestimated awkwardness, and a marginally significant conversation \times measurement phase interaction effect, $F(1, 101) = 3.73, p = .056, \eta_p^2 = .04$, such that participants overestimated awkwardness more for deep questions than for control questions. Participants in the deep condition expected their conversations to be more awkward than participants in the control condition, $t(101) = -2.63, p = .010, 95\% \text{ CI}_{\text{difference}} = [-2.01, -0.28], d = -0.52$, but experiences did not differ, $t(101) = -1.04, p = .300, 95\% \text{ CI}_{\text{difference}} = [-1.19, 0.37], d = -0.21$. Participants overestimated awkwardness in both the deep condition, *paired* $t(51) = 6.66, p < .001, 95\% \text{ CI}_{\text{difference}} = [1.26, 2.35], d = 0.92$, and the control condition, *paired* $t(50) = 4.04, p < .001, 95\% \text{ CI}_{\text{difference}} = [0.54, 1.61], d = 0.57$.

Connectedness. We observed a main effect of conversation, $F(1, 101) = 10.42, p = .002, \eta_p^2 = .09$, such that predicted or actual connectedness was greater for participants in the deep condition, and a main effect of measurement phase, $F(1, 101) = 42.11, p < .001, \eta_p^2 = .29$, such that participants underestimated connectedness. The conversation \times measurement phase interaction effect was non-significant, $F(1, 101) = 0.02, p = .895, \eta_p^2 = .0002$. Participants in the deep condition predicted greater connectedness than participants in the control condition, $t(101) = -3.32, p = .001, 95\% \text{ CI}_{\text{difference}} = [-1.60, -0.40], d = -0.65$, and also reported greater experiences of connectedness, $t(101) = -2.63, p = .010, 95\% \text{ CI}_{\text{difference}} = [-1.83, -0.26], d = -0.52$. Participants underestimated connectedness in both the control condition, *paired* $t(50) = -4.67, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.41, -0.56], d = -0.65$, and the deep condition, *paired* $t(51) = -4.52, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.48, -0.57], d = -0.63$.

Happiness. The main effect of conversation was non-significant, $F(1, 101) = 2.54, p = .114, \eta_p^2 = .02$, but we observed a significant main effect of measurement phase, $F(1, 101) = 43.60, p < .001, \eta_p^2 = .30$, such that participants underestimated happiness. The conversation \times measurement phase interaction effect was non-significant, $F(1, 101) = 2.65, p = .107, \eta_p^2 = .03$. Predicted happiness did not differ by condition, $t(101) = -0.72, p = .472, 95\% \text{ CI}_{\text{difference}} = [-0.83, 0.38], d = -0.14$, but experienced happiness was marginally greater in the deep condition, $t(101) = -1.96, p = .052, 95\% \text{ CI}_{\text{difference}} = [-1.51, 0.007], d = -0.39$. Participants in the control condition underestimated happiness, *paired* $t(50) = -3.51, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.28, -0.35], d = -0.49$, as did participants in the deep condition, *paired* $t(51) = -5.83, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.81, -0.88], d = -0.81$.

Gender Effects

Neither the amount of miscalibration on the primary measures, nor differences in miscalibration between the control and deep conditions, differed significantly between same-gender and mixed-gender pairs, $F_s(1, 96) \leq 3.04$, $p_s \geq .080$, $\eta_p^2_s \leq .03$.

Experiment 4a

Full Output from ANOVAs in Main Text

A 2 (conversation: shallow, deep) \times 2 (target: self, partner) ANOVA on participants' care expectations yielded a significant main effect of conversation, $F(1, 101) = 38.79, p < .001, \eta_p^2 = .28$, indicating that expected care was higher for deep conversations, a main effect of target, $F(1, 101) = 32.90, p < .001, \eta_p^2 = .25$, indicating that participants thought they would care more about their own responses than their partner would, and a significant interaction, $F(1, 101) = 6.00, p = .016, \eta_p^2 = .06$, indicating that the self/other caring gap was significantly larger in the deep condition than in the shallow condition.

A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA on the partner care measure with repeated measures on the second factor produced a significant main effect of conversation, $F(1, 101) = 16.24, p < .001, \eta_p^2 = .14$, such that more care was expected or experienced in the deep condition than the shallow condition, and a significant main effect of measurement phase, $F(1, 101) = 93.22, p < .001, \eta_p^2 = .48$, such that participants underestimated partner care. The conversation \times measurement phase interaction effect was non-significant, $F(1, 101) = 0.29, p = .591, \eta_p^2 = .003$.

For awkwardness, a 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the second factor yielded a significant main effect of conversation, $F(1, 101) = 39.70, p < .001, \eta_p^2 = .28$, indicating that participants in the deep condition reported greater awkwardness, and a main effect of measurement phase, $F(1, 101) = 52.65, p < .001, \eta_p^2 = .34$, indicating that participants overestimated awkwardness across conditions. These main effects were qualified by a significant conversation \times measurement phase interaction, $F(1, 101) = 4.30, p = .041, \eta_p^2 = .04$, indicating

that participants in the deep condition overestimated awkwardness more than participants in the shallow condition.

For connectedness, a 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the second factor yielded a main effect of conversation, $F(1, 101) = 25.90, p < .001, \eta_p^2 = .20$, indicating that participants in the deep condition reported greater connectedness than participants in the shallow condition, and a main effect of measurement phase, $F(1, 101) = 144.32, p < .001, \eta_p^2 = .59$, indicating that participants underestimated connectedness. These main effects were qualified by a significant conversation \times measurement phase interaction effect, $F(1, 101) = 11.76, p < .001, \eta_p^2 = .10$, indicating that participants in the shallow condition underestimated connectedness significantly more than participants in the deep condition.

For happiness, a 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the second factor yielded a main effect of conversation, $F(1, 101) = 5.35, p = .023, \eta_p^2 = .05$, indicating that participants in the deep condition reported greater happiness than participants in the shallow condition, and a main effect of measurement phase, $F(1, 101) = 130.18, p < .001, \eta_p^2 = .56$, indicating that participants generally underestimated their happiness. These main effects were qualified by a significant conversation \times measurement phase interaction effect, $F(1, 101) = 13.18, p < .001, \eta_p^2 = .12$, indicating that participants underestimated happiness significantly more in the shallow condition than the deep condition.

Mediational Analyses at the Level of the Individual

We conducted mediational analyses at the level of the individual, with measurement phase (expectations vs. experiences) as the independent variable, partner care as the mediating

variable, and the primary measures as dependent variables, separately for the shallow and deep conditions. In the shallow condition, the indirect effects were significant for awkwardness ($b = 0.38$, 95% CI = [0.03, 0.72]), connectedness ($b = -0.46$, 95% CI = [-0.82, -0.20]), and happiness ($b = -0.41$, 95% CI = [-0.83, -0.13]). In the deep condition, the indirect effects were non-significant for awkwardness ($b = 0.18$, 95% CI = [-0.25, 0.65]), significant for connectedness ($b = -0.41$, 95% CI = [-0.74, -0.18]), and significant for happiness ($b = -0.41$, 95% CI = [-0.75, -0.11]).

Analyses Without Exclusions

We re-analyzed the data with all 105 pairs of participants included.

Care. A 2 (conversation: shallow, deep) \times 2 (target: self, partner) ANOVA on participants' expectations yielded a main effect of conversation, $F(1, 103) = 34.89$, $p < .001$, $\eta_p^2 = .25$, a main effect of target, $F(1, 103) = 33.99$, $p < .001$, $\eta_p^2 = .25$, and a significant interaction, $F(1, 103) = 4.63$, $p = .034$, $\eta_p^2 = .04$. Although participants in the shallow condition expected to care significantly more about their responses to the questions ($M = 5.68$, $SD = 1.57$) than their partner would ($M = 5.02$, $SD = 1.67$), *paired* $t(52) = -2.83$, $p = .007$, 95% CI_{difference} = [-1.13, -0.19], $d = -0.39$, this gap was significantly larger among participants in the deep condition ($M_{\text{self}} = 7.45$, $SD_{\text{self}} = 1.35$; $M_{\text{partner}} = 6.02$, $SD_{\text{partner}} = 1.44$), *paired* $t(51) = -5.24$, $p < .001$, 95% CI_{difference} = [-1.98, -0.88], $d = -0.73$.

A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA on the partner care measure with repeated measures on the second factor produced a significant main effect of conversation, $F(1, 103) = 16.88$, $p < .001$, $\eta_p^2 = .14$, such that more care was expected or experienced in the deep condition than the shallow condition, and a significant main effect of measurement phase, $F(1, 103) = 98.24$, $p < .001$, $\eta_p^2 = .49$, such that

participants underestimated partner care. The conversation \times measurement phase interaction effect was non-significant, $F(1, 103) = 0.18, p = .672, \eta_p^2 = .002$, indicating that participants underestimated the recipients' care similarly in the deep and shallow conditions.

Awkwardness. A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA, with repeated measures on the second factor, yielded main effect of conversation, $F(1, 103) = 35.60, p < .001, \eta_p^2 = .26$, a main effect of measurement phase, $F(1, 103) = 49.96, p < .001, \eta_p^2 = .33$, and a conversation \times measurement phase interaction effect, $F(1, 103) = 4.55, p = .035, \eta_p^2 = .04$. Participants in the shallow condition overestimated awkwardness, *paired* $t(52) = 3.55, p < .001, 95\% \text{ CI}_{\text{difference}} = [0.46, 1.64], d = 0.49$, but participants in the deep condition overestimated awkwardness to a greater extent, *paired* $t(51) = 6.39, p < .001, 95\% \text{ CI}_{\text{difference}} = [1.34, 2.57], d = 0.89$.

Connectedness. We observed a main effect of conversation, $F(1, 103) = 26.86, p < .001, \eta_p^2 = .21$, a main effect of measurement phase, $F(1, 103) = 148.60, p < .001, \eta_p^2 = .59$, and a conversation \times measurement phase interaction effect $F(1, 103) = 11.59, p < .001, \eta_p^2 = .10$. Participants in the deep condition underestimated connectedness, *paired* $t(51) = -7.02, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.49, -0.83], d = -0.97$, but participants in the shallow condition underestimated connectedness to a greater degree, *paired* $t(52) = -10.03, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.47, -1.65], d = -1.38$. Participants in the deep condition predicted greater connectedness than participants in the shallow condition, $t(103) = 5.48, p < .001, 95\% \text{ CI}_{\text{difference}} = [0.88, 1.89], d = 1.07$, and experienced greater connectedness, $t(103) = 2.56, p = .012, 95\% \text{ CI}_{\text{difference}} = [0.11, 0.86], d = 0.50$.

Happiness. We observed a main effect of conversation, $F(1, 103) = 4.62, p = .034, \eta_p^2 = .04$, and a main effect of measurement phase, $F(1, 103) = 133.98, p < .001, \eta_p^2 = .57$, qualified

by a conversation \times measurement phase interaction effect, $F(1, 103) = 12.52, p < .001, \eta_p^2 = .11$. Participants in the shallow condition underestimated happiness, *paired* $t(52) = -9.86, p < .001$, 95% $CI_{\text{difference}} = [-2.35, -1.56], d = -1.35$, and they did so to a greater degree than participants in the deep condition, *paired* $t(51) = -6.28, p < .001$, 95% $CI_{\text{difference}} = [-1.37, -0.71], d = -0.87$. Participants in the deep condition predicted greater happiness, $t(103) = 3.37, p = .001$, 95% $CI_{\text{difference}} = [0.36, 1.41], d = 0.66$, but happiness experiences did not differ, $t(103) = -0.14, p = .890$, 95% $CI_{\text{difference}} = [-0.44, 0.39], d = -0.03$.

Mediational analyses. In the shallow condition, the indirect effects were non-significant for awkwardness ($b = 0.19$, 95% $CI = [-0.25, 0.65]$), non-significant for connectedness ($b = -0.44$, 95% $CI = [-1.03, 0.03]$), and non-significant for happiness ($b = -0.43$, 95% $CI = [-0.93, 0.002]$). In the deep condition, the indirect effects were non-significant for awkwardness ($b = 0.05$, 95% $CI = [-0.41, 0.55]$), significant for connectedness ($b = -0.47$, 95% $CI = [-0.86, -0.17]$), and non-significant for happiness ($b = -0.35$, 95% $CI = [-0.68, 0.01]$).

Desire for shallow and deep conversation. Participants reported that they wanted to engage in more small talk than they typically do ($M = 1.10, SD = 2.34$), $t(209) = 6.82, p < .001$, 95% $CI = [0.78, 1.42], d = 0.47$, and more deep conversation ($M = 1.41, SD = 2.17$), $t(209) = 9.42, p < .001$, 95% $CI = [1.11, 1.70], d = 0.65$. Differences between small talk and deep conversation were not significant, *paired* $t(209) = -1.64, p = .102$, 95% $CI_{\text{difference}} = [-0.68, 0.06], d = -0.11$.

Gender Effects

Neither the amount of miscalibration on the primary measures, nor differences in miscalibration between the shallow and deep conditions, differed significantly between same-gender and mixed-gender pairs, $F_s(1, 99) \leq 1.53, p_s \geq .219, \eta_p^2_s \leq .02$.

Experiment 4b

Method

Participants. We recruited one hundred and two pairs of participants from a master's degree student orientation event ($N = 202$ individuals after exclusions; $M_{\text{age}} = 36.63$; $SD_{\text{age}} = 5.10$; 19.31% female; 37.13% Caucasian). We excluded 1 pair from analyses because the participants did not discuss the conversation questions they were assigned. In addition, another 21 participants could not be matched with their partner after the session because either 1 or 3 participants reported the same pair ID; we did not analyze data from these participants.

Procedure. The procedure was identical to Experiment 4a with several exceptions. First, rather than manipulate the intimacy of the discussion questions, we manipulated the intimacy of the conversation that participants were asked to have and then provided several shallow or deep discussion questions as examples. Participants were not required to restrict their conversations to the questions we provided. Participants in the shallow condition read the following instructions:

After being paired up, we would like you to have a light and easy conversation with the other person. You should try to maintain small talk and focus on relatively light issues that occur to you. Try to guide the conversation toward topics that are easy to talk about and that are not too personally intimate or emotional. Please respect the other person's privacy. You are strangers at this point, after all. Both you and the person you will speak with are receiving this instruction to keep the conversation light and easy. Here are some examples of questions you should discuss. Please restrict your conversation to relatively light and easy topics like these:

- How is your day going so far?
- Have you been to the Harper Center before?

- What do you think about the weather today?
- Can you describe a conversation you had with another person earlier today?
- What do you think of [removed for blinded review]?

Participants in the deep condition read the following instructions:

After being paired up, we would like you to have a deep and meaningful conversation with the other person. You should try to go beyond the surface, beyond small talk, by revealing deep and personal information about yourself. Try to guide the conversation toward topics that are really important and that are personally intimate and emotional. Please try to have a deep and meaningful interaction with this person. You are part of the same community, after all. Both you and the person you will speak with are receiving this instruction to have a deep and meaningful interaction. Here are some examples of questions you should discuss. Please restrict your conversation to relatively deep and meaningful topics like these:

- If I was to become a good friend of yours, what would be most important for me to know about you?
- For what in your life do you feel most grateful? Please tell me about it.
- If you could undo one mistake you have made in your life, what would it be and why would you undo it?
- Can you tell me about one of the last times you cried in front of another person?
- What are you most afraid of starting this program at [removed for blinded review]?

Second, we modified the happiness measure by asking participants to report their overall happiness (0 = *not at all happy*; 10 = *extremely happy*) rather than how happy they felt about the conversation. Before the intimacy manipulation, participants reported how happy they felt at the moment to establish a baseline measure, and then later predicted how happy they would feel following the conversation and reported how happy they actually felt after the conversation.

Third, after the conversation, participants completed a manipulation check by reporting how intimate their conversation was (0 = *light conversation and small talk*; 10 = *meaningful conversation and deep talk*). Fourth, participants also reported how intimate their typical conversations with strangers were, and how intimate they wished their conversations with strangers would be, using the same scale.

Fifth, after completing these items, participants reported how positive or negative their conversation was (-5 = *very negative*; 0 = *equally positive and negative*; 5 = *very positive*) and then reported how the conversation made them feel in free-response format. These free-responses were used only for illustrative purposes and are not analyzed quantitatively.

Finally, participants reported demographic information and were thanked and debriefed.

Results (Experiment 4b)

As in Experiment 4a, the strength-of-bond and liking measures were highly correlated in expectations ($\alpha = .79$) and in experiences ($\alpha = .83$) and so we combined these measures to form a connectedness scale.

Manipulation check. The conversation intimacy manipulation was effective: Participants in the deep condition ($M = 7.29$, $SD = 1.54$) reported deeper and more meaningful conversations than participants in the shallow condition ($M = 5.72$, $SD = 1.86$), $t(99) = 4.64$, $p < .001$, 95% $CI_{\text{difference}} = [0.90, 2.25]$, $d = 0.92$.

Care measures. As predicted, participants expected to care more about their own responses than their partner would, and this self/other caring gap was significantly larger in the deep condition than in the shallow condition. A 2 (conversation: shallow, deep) \times 2 (target: self, partner) ANOVA on participants' expectations yielded a significant main effect of conversation, $F(1, 99) = 8.11, p = .005, \eta_p^2 = .08$, indicating that expected care was higher for deep conversations, a main effect of target, $F(1, 99) = 48.33, p < .001, \eta_p^2 = .33$, indicating that participants thought they would care more about their own responses than their partner would, and a significant interaction, $F(1, 99) = 5.17, p = .025, \eta_p^2 = .05$. Although participants in the shallow condition expected to care significantly more about their responses to the questions ($M = 6.54, SD = 1.42$) than their partner would ($M = 5.95, SD = 1.48$), $paired\ t(50) = -4.49, p < .001, 95\% CI_{\text{difference}} = [-0.85, -0.32], d = -0.63$, this gap was significantly larger among participants in the deep condition ($M_s = 7.59$ vs. 6.43 , respectively; $SDs = 1.47$ vs. 1.59), $paired\ t(49) = -5.37, p < .001, 95\% CI_{\text{difference}} = [-1.59, -0.73], d = -0.76$.

Perhaps more important, comparisons with actual experience indicated that participants systematically underestimated how much they would actually perceive their partner to care about their responses. A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA on the partner care measure with repeated measures on the second factor produced a non-significant main effect of conversation, $F(1, 99) = 1.28, p = .261, \eta_p^2 = .01$, but a significant main effect of measurement phase, $F(1, 99) = 83.14, p < .001, \eta_p^2 = .46$, such that participants underestimated partner care. The conversation \times measurement phase interaction effect was non-significant, $F(1, 99) = 1.66, p = .201, \eta_p^2 = .02$, indicating that participants underestimated the recipients' care similarly in the deep and shallow conditions. Consistent with

our hypotheses, participants underestimated their partner's care significantly more than they underestimated their own, $F(1, 99) = 37.86, p < .001, \eta_p^2 = .28$.

Awkwardness. Participants again overestimated how awkward the conversation would be across conditions, and did so somewhat more in the deep condition than in the shallow condition. A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the second factor yielded a significant main effect of conversation, $F(1, 99) = 18.83, p < .001, \eta_p^2 = .16$, indicating that participants in the deep condition reported greater awkwardness, and a main effect of measurement phase, $F(1, 99) = 84.92, p < .001, \eta_p^2 = .46$, indicating that participants overestimated awkwardness across conditions. These main effects were qualified by a marginally significant conversation \times measurement phase interaction effect, $F(1, 99) = 3.69, p = .058, \eta_p^2 = .04$. Although participants overestimated how awkward the conversation would be in the shallow condition, *paired t*(50) = 6.28, $p < .001$, 95% $CI_{\text{difference}} = [1.13, 2.19]$, $d = 0.88$, participants overestimated feelings of awkwardness marginally more in the deep condition, *paired t*(49) = 6.80, $p < .001$, 95% $CI_{\text{difference}} = [1.78, 3.28]$, $d = 0.96$.

Connectedness. Participants again underestimated connectedness across conditions, but unexpectedly, participants in the shallow condition underestimated connectedness more than participants in the deep condition. A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the second factor yielded a significant main effect of conversation, $F(1, 99) = 7.86, p = .006, \eta_p^2 = .07$, indicating that connectedness was greater for participants in the deep condition, and a significant main effect of measurement phase, $F(1, 99) = 85.38, p < .001, \eta_p^2 = .46$, indicating that participants underestimated connectedness. These main effects were qualified by a significant conversation \times

measurement phase interaction effect, $F(1, 99) = 17.96, p < .001, \eta_p^2 = .15$. Although participants underestimated how connected they would feel in the deep condition, *paired* $t(49) = -3.37, p = .001, 95\% \text{ CI}_{\text{difference}} = [-1.05, -0.26], d = -0.48$, they did so significantly more in the shallow condition, *paired* $t(50) = -10.05, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.12, -1.41], d = -1.41$. Participants in the deep condition expected to feel more connected than did participants in the shallow condition, $t(99) = 4.91, p < .001, 95\% \text{ CI}_{\text{difference}} = [0.66, 1.54], d = 0.98$, but experiences of connectedness did not differ significantly between conditions, $t(99) = -0.04, p = .970, 95\% \text{ CI}_{\text{difference}} = [-0.49, 0.48], d = -0.01$.

Happiness. Participants underestimated happiness across conditions, but participants in the shallow condition underestimated how happy they would feel somewhat more than participants in the deep condition. A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the second factor yielded a non-significant main effect of conversation, $F(1, 99) = 0.02, p = .898, \eta_p^2 = .0002$, but a significant main effect of measurement phase, $F(1, 99) = 64.36, p < .001, \eta_p^2 = .39$, indicating that participants generally underestimated their happiness. This main effect was qualified by a marginally significant conversation \times measurement phase interaction effect, $F(1, 99) = 3.21, p = .076, \eta_p^2 = .03$. Although participants significantly underestimated how happy they would feel about the conversation in the deep condition, *paired* $t(49) = -3.87, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.00, -0.32], d = -0.55$, they did so marginally more in the shallow condition, *paired* $t(50) = -8.20, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.29, -0.78], d = -1.15$. Participants in the deep condition did not expect to feel significantly happier about their conversations than participants in the shallow condition, $t(99) = 0.70, p = .486, 95\% \text{ CI}_{\text{difference}} = [-0.30, 0.63], d = 0.14$, nor did they experience

greater happiness after speaking, $t(99) = -1.02$, $p = .309$, 95% $CI_{\text{difference}} = [-0.63, 0.20]$, $d = -0.20$.

In both the shallow and deep conditions, speaking with the other person enhanced participants' mood compared to how they felt before the conversation. Participants in the shallow condition felt happier after their conversations ($M = 8.25$, $SD = 0.83$) than they did at baseline ($M = 7.62$, $SD = 0.91$), *paired* $t(50) = -6.08$, $p < .001$, 95% $CI_{\text{difference}} = [-0.85, -0.43]$, $d = -0.85$. Participants in the deep condition also felt happier after their conversations ($M = 8.04$, $SD = 7.28$) than they did at baseline ($M = 7.28$, $SD = 1.08$), *paired* $t(49) = -5.16$, $p < .001$, 95% $CI_{\text{difference}} = [-1.06, -0.46]$, $d = -0.73$. Changes in happiness did not differ significantly by condition, $F(1, 99) = 0.46$, $p = .497$, $\eta_p^2 = .005$. This suggests that connecting with strangers, whether in a relatively shallow or deeper way, could enhance people's mood to a degree that would surprise those engaging in the conversation (see also Epley & Schroeder, 2014).

Mediational analyses. We found some evidence that underestimating a partner's care could explain why participants underestimated how connected and happy they would feel following the conversation. We performed a series of within-pairs mediational analyses using measurement phase (expectation vs. experience) as the independent variable and partner care as the mediating variable. In the shallow condition, the indirect effects were non-significant for awkwardness ($b = 0.35$, 95% $CI = [-0.31, 0.99]$), significant for connectedness ($b = -0.54$, 95% $CI = [-1.02, -0.19]$), and significant for happiness ($b = -0.36$, 95% $CI = [-0.63, -0.17]$). In the deep condition, the indirect effects were also non-significant for awkwardness ($b = 0.37$, 95% $CI = [-0.25, 0.92]$), significant for connectedness ($b = -0.53$, 95% $CI = [-0.93, -0.24]$), and significant for happiness ($b = -0.39$, 95% $CI = [-0.89, -0.06]$). These findings suggest that underestimating others' care may help to explain why deep conversations between strangers

unfold more positively than people expect, and why even shallow conversations may turn out more positively than people expect as well.

Desired conversation intimacy. Unlike in Experiment 4a, Participants in Experiment 4b reported how intimate their normal conversations with strangers were, and how intimate they wished them to be, on a scale ranging from 0 (*light conversation and small talk*) to 10 (*meaningful conversation and deep talk*). Again indicating a general preference for deeper conversation in daily life, participants reported that their typical conversations with strangers ($M = 3.87$, $SD = 2.23$) were less deep and meaningful than they wished they would be ($M = 5.82$, $SD = 2.43$), $paired\ t(201) = -11.93$, $p < .001$, $95\% CI_{\text{difference}} = [-2.27, -1.63]$, $d = -0.84$. These results again suggest that a reluctance to engage in deeper conversation does not stem from people's own disinterest in doing so, but rather may stem from some of the miscalibrated expectations about others' interest and care described earlier.

Mediational Analyses at the Level of the Individual

We conducted mediational analyses at the level of the individual, with measurement phase (expectations vs. experiences) as the independent variable, partner care as the mediating variable, and the primary measures as dependent variables, separately for the shallow and deep conditions. In the shallow condition, the indirect effects were non-significant for awkwardness ($b = 0.21$, $95\% CI = [-0.15, 0.61]$), significant for connectedness ($b = -0.41$, $95\% CI = [-0.75, -0.18]$), and significant for happiness ($b = -0.25$, $95\% CI = [-0.47, -0.10]$). In the deep condition, the indirect effects were also non-significant for awkwardness ($b = 0.33$, $95\% CI = [-0.01, 0.74]$), significant for connectedness ($b = -0.43$, $95\% CI = [-0.76, -0.21]$), and significant for happiness ($b = -0.38$, $95\% CI = [-0.72, -0.16]$).

Analyses Without Exclusions

We re-analyzed data with all 102 pairs included.

Care. A 2 (conversation: shallow, deep) \times 2 (target: self, partner) ANOVA on participants' expectations yielded a significant main effect of conversation, $F(1, 100) = 8.64, p = .004, \eta_p^2 = .08$, a main effect of target, $F(1, 100) = 50.24, p < .001, \eta_p^2 = .33$, and a significant interaction, $F(1, 100) = 4.51, p = .036, \eta_p^2 = .04$. Although participants in the shallow condition expected to care significantly more about their responses to the questions ($M = 6.54, SD = 1.41$) than their partner would ($M = 5.91, SD = 1.49$), *paired t*(51) = 4.67, $p < .001$, 95% $CI_{\text{difference}} = [0.36, 0.89]$, $d = 0.65$, this gap was significantly larger among participants in the deep condition ($M_{\text{self}} = 7.59, SD_{\text{self}} = 1.47$; $M_{\text{partner}} = 6.43, SD_{\text{partner}} = 1.59$), *paired t*(49) = 5.37, $p < .001$, 95% $CI_{\text{difference}} = [0.73, 1.59]$, $d = 0.76$.

A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA on the partner care measure with repeated measures on the second factor produced a non-significant main effect of conversation, $F(1, 100) = 1.39, p = .241, \eta_p^2 = .01$, but a significant main effect of measurement phase, $F(1, 100) = 85.24, p < .001, \eta_p^2 = .46$, such that participants underestimated partner care. The conversation \times measurement phase interaction effect was non-significant, $F(1, 100) = 2.12, p = .148, \eta_p^2 = .02$, indicating that participants underestimated the recipients' care similarly in the deep and shallow conditions.

Awkwardness. A 2 (conversation: shallow, deep) \times 2 (measurement phase: expectations, experiences) ANOVA, with repeated measures on the second factor, yielded a main effect of conversation, $F(1, 100) = 19.54, p < .001, \eta_p^2 = .16$, a main effect of measurement phase, $F(1, 100) = 87.16, p < .001, \eta_p^2 = .47$, and a marginally significant conversation \times measurement phase interaction effect, $F(1, 100) = 3.62, p = .060, \eta_p^2 = .03$. Participants in the shallow condition overestimated awkwardness, *paired t*(51) = 6.46, $p < .001$, 95% $CI_{\text{difference}} = [1.15, 2.19]$, $d =$

0.90, but participants in the deep condition did so to a marginally greater extent, *paired* $t(49) = 6.80, p < .001, 95\% \text{ CI}_{\text{difference}} = [1.78, 3.28], d = 0.96$.

Connectedness. We observed a main effect of conversation, $F(1, 100) = 8.11, p = .005, \eta_p^2 = .07$, a main effect of measurement phase, $F(1, 100) = 87.97, p < .001, \eta_p^2 = .47$, and a conversation \times measurement phase interaction effect, $F(1, 100) = 19.10, p < .001, \eta_p^2 = .16$. Participants in the shallow condition underestimated connectedness, *paired* $t(51) = -10.25, p < .001, 95\% \text{ CI}_{\text{difference}} = [-2.15, -1.45], d = -1.42$, as did participants in the deep condition, *paired* $t(49) = -3.37, p = .001, 95\% \text{ CI}_{\text{difference}} = [-1.05, -0.26], d = -0.48$. Participants in the deep condition predicted greater connectedness than participants in the shallow condition, $t(100) = 5.03, p < .001, 95\% \text{ CI}_{\text{difference}} = [0.68, 1.56], d = 1.00$, but experienced connectedness did not differ across conditions, $t(100) = -0.09, p = .925, 95\% \text{ CI}_{\text{difference}} = [-0.50, 0.46], d = -0.02$.

Happiness. The main effect of conversation was non-significant, $F(1, 100) = 0.001, p = .974, \eta_p^2 = .00001$, but we observed a main effect of measurement phase, $F(1, 100) = 66.19, p < .001, \eta_p^2 = .40$, and a marginally significant conversation \times measurement phase interaction effect, $F(1, 100) = 3.42, p = .067, \eta_p^2 = .03$. Participants in the shallow condition underestimated happiness, *paired* $t(51) = -8.42, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.30, -0.80], d = -1.17$, as did participants in the deep condition, *paired* $t(49) = -3.87, p < .001, 95\% \text{ CI}_{\text{difference}} = [-1.00, -0.32], d = -0.55$. Happiness expectations did not differ by condition, $t(100) = 0.80, p = .424, 95\% \text{ CI}_{\text{difference}} = [-0.28, 0.65], d = 0.16$, nor did happiness experiences, $t(100) = -0.96, p = .340, 95\% \text{ CI}_{\text{difference}} = [-0.61, 0.21], d = -0.19$.

Mediational analyses. In the shallow condition, the indirect effects were non-significant for awkwardness ($b = 0.35, 95\% \text{ CI} = [-0.31, 1.04]$), significant for connectedness ($b = -0.54, 95\% \text{ CI} = [-1.03, -0.20]$), and significant for happiness ($b = -0.36, 95\% \text{ CI} = [-0.62, -0.16]$). In

the deep condition, the indirect effects were also non-significant for awkwardness ($b = 0.36$, 95% CI = [-0.27, 0.89]), significant for connectedness ($b = -0.60$, 95% CI = [-0.99, -0.25]), and significant for happiness ($b = -0.41$, 95% CI = [-0.82, -0.04]).

Desire for shallow and deep conversation. Participants reported that their typical conversations with strangers were less deep and meaningful than they wished they would be, *paired t*(203) = -11.94, $p < .001$, 95% CI_{difference} = [-2.28, -1.63], $d = -0.84$.

Gender Effects

Neither the amount of miscalibration on the primary measures, nor differences in miscalibration between the shallow and deep conditions, differed significantly between same-gender and mixed-gender pairs, $F_s(1, 97) \leq 3.63$, $p_s \geq .060$, $\eta_p^2_s \leq .04$. The one exception was that same-gender pairs were significantly more likely than mixed-gender pairs to underestimate their connectedness, $F(1, 97) = 4.52$, $p = .036$, $\eta_p^2 = .04$.

Experiment 5

Conversation Manipulations

First, participants generated two control questions. The wording of the instructions was tailored to each relationship context. Participants were told:

We would like you to begin by generating two questions. These should be the types of questions that you would typically ask while [getting to know another person for the first time] [spending time with the other person]. Please select questions that you would actually be willing to ask and answer later in this study, and these questions should be the types of questions that you would typically ask while [getting to know somebody new] [speaking with this person]. Note that these should be questions you're able to both ask the other person and answer yourself.

Participants then generated two deep questions. Specifically, they were told:

Next we would like you to generate two more questions. This time, please generate two questions that are deeper and more intimate than the types of questions that you would typically ask while [getting to know another person for the first time] [speaking with this person]. In other words, we would like you to generate questions that go beyond the surface, beyond small talk, to probe deeper subject matter that might be more personal or emotional. For example, you might ask the person about important experiences they've had or activities they've enjoyed. You might ask the person to reveal something important about them. You might ask them what they feel grateful for or to describe something they wish they'd done differently in their lives. The questions you generate should require both you and your partner to reveal something about yourselves that you

might not normally reveal while [first getting to know one another] [speaking with one another]. These should be topics that are more intimate and go deeper than the ones you wrote down in the previous set of questions. Again, note that these should be questions you're able to both ask the other person and answer yourself.

Full Output from ANOVAs in Main Text

The 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (target: self, partner) ANOVA on participants' expectations yielded a significant main effect of target, $F(1, 196) = 6.73, p = .010, \eta_p^2 = .03$, indicating that participants thought they would care more about their own responses than their partner would, a main effect of conversation, $F(1, 196) = 23.39, p < .001, \eta_p^2 = .11$, indicating that participants in the deep conditions predicted greater partner care, and a main effect of relationship, $F(1, 196) = 61.69, p < .001, \eta_p^2 = .24$, indicating the participants in the close conditions predicted greater partner care. These main effects were qualified by a significant conversation \times target interaction, $F(1, 196) = 8.86, p = .003, \eta_p^2 = .04$, indicating that participants in the deep conditions predicted greater self/other differences in care, and a significant relationship \times target interaction, $F(1, 196) = 8.86, p = .003, \eta_p^2 = .04$, indicating that participants in the distant conditions predicted greater self/other differences in care.

A 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the third factor and partner care as the dependent measure produced a significant main effect of conversation, $F(1, 196) = 13.78, p < .001, \eta_p^2 = .07$, such that partner care was greater among deep pairs, a main effect of relationship, $F(1, 196) = 74.84, p < .001, \eta_p^2 = .28$, such that close pairs reported greater partner care than distant pairs, a main effect of measurement phase, $F(1, 196) = 106.68, p < .001, \eta_p^2 = .35$, such that participants underestimated partner care, and a relationship \times measurement

phase interaction effect, $F(1, 196) = 55.84, p < .001, \eta_p^2 = .22$, such that distant pairs underestimated partner care more than close pairs.

For awkwardness, a 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the third factor produced a main effect of conversation, $F(1, 196) = 6.59, p = .011, \eta_p^2 = .03$, such that awkwardness was greater for participants in the deep condition than participants in the control condition, a main effect of measurement phase, $F(1, 196) = 92.79, p < .001, \eta_p^2 = .32$, such that participants overestimated awkwardness, a conversation \times measurement phase interaction, $F(1, 196) = 4.94, p = .027, \eta_p^2 = .02$, such that participants in the deep condition overestimated awkwardness more than participants in the control condition, and a relationship \times measurement phase interaction, $F(1, 196) = 11.05, p = .001, \eta_p^2 = .05$, such that distant pairs overestimated awkwardness more than close pairs.

For enjoyment, a 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the third factor produced a significant effect of relationship, $F(1, 196) = 16.72, p < .001, \eta_p^2 = .08$, such that close pairs reported greater enjoyment than distant pairs, an effect of measurement phase, $F(1, 196) = 135.31, p < .001, \eta_p^2 = .41$, such that participants underestimated enjoyment, and a significant relationship \times measurement phase interaction, $F(1, 196) = 29.58, p < .001, \eta_p^2 = .13$, such that distant pairs underestimated enjoyment more than close pairs.

For happiness, a 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVA with repeated measures on the third factor produced a main effect of relationship, $F(1, 196) = 22.87, p < .001, \eta_p^2 = .10$, such that close pairs reported greater happiness than distant pairs, a main effect of measurement phase,

$F(1, 196) = 188.47, p < .001, \eta_p^2 = .49$, such that participants underestimated happiness, and a relationship \times measurement phase interaction effect, $F(1, 196) = 29.96, p < .001, \eta_p^2 = .13$, such that distant pairs underestimated happiness more than close pairs.

Mediational Analyses at the Level of the Individual

We conducted mediational analyses at the level of the individual, with measurement phase (expectations vs. experiences) as the independent variable, partner care as the mediating variable, and the primary measures as dependent variables, separately for each condition. Among participants in the control condition who spoke with distant strangers, the indirect effects were non-significant for awkwardness ($b = 0.21, 95\% \text{ CI} = [-0.14, 0.60]$), non-significant for enjoyment ($b = -0.25, 95\% \text{ CI} = [-0.63, 0.13]$), and significant for happiness ($b = -0.46, 95\% \text{ CI} = [-0.74, -0.17]$). Among participants in the control condition who spoke with close others, the indirect effects were non-significant for awkwardness ($b = 0.02, 95\% \text{ CI} = [-0.02, 0.15]$), non-significant for enjoyment ($b = -0.06, 95\% \text{ CI} = [-0.26, 0.06]$), and non-significant for happiness ($b = -0.06, 95\% \text{ CI} = [-0.23, 0.07]$).

Among participants in the deep condition who spoke with distant strangers, the indirect effects were significant for awkwardness ($b = 0.61, 95\% \text{ CI} = [0.21, 1.13]$), non-significant for enjoyment ($b = -0.04, 95\% \text{ CI} = [-0.67, 0.55]$), and significant for happiness ($b = -0.44, 95\% \text{ CI} = [-0.81, -0.18]$). Among participants in the deep condition who spoke with close others, the indirect effects were non-significant for awkwardness ($b = 0.09, 95\% \text{ CI} = [-0.03, 0.32]$), significant for enjoyment ($b = -0.17, 95\% \text{ CI} = [-0.36, -0.04]$), and significant for happiness ($b = -0.11, 95\% \text{ CI} = [-0.23, -0.04]$).

We then conducted between-individuals mediational analyses with relationship (close vs. distant) as the independent variable, underestimation of partner care as the mediating variable,

and the amount of miscalibration for each of the primary measures as dependent variables in separate analyses. Among participants in the control conditions, the indirect effects were non-significant for awkwardness ($b = 0.22$, 95% CI = [-0.04, 0.52]), significant for enjoyment ($b = -0.42$, 95% CI = [-0.69, -0.16]), and significant for happiness ($b = -0.53$, 95% CI = [-0.80, -0.31]). Among participants in the deep conditions, the indirect effects were significant for awkwardness ($b = 0.36$, 95% CI = [0.11, 0.70]), non-significant for enjoyment ($b = -0.23$, 95% CI = [-0.60, 0.09]), and significant for happiness ($b = -0.33$, 95% CI = [-0.61, -0.13]).

Analyses Without Exclusions

We reanalyzed the data with all 204 pairs included.

Care. A 2 (conversation: control, deep) \times 2 (Relationship: distant, close) \times 2 (target: self, partner) ANOVA on participants' expectations yielded a significant main effect of target, $F(1, 200) = 6.36$, $p = .012$, $\eta_p^2 = .03$, indicating that participants thought they would care more about their own responses than their partner would, qualified by a significant conversation \times target interaction, $F(1, 200) = 8.56$, $p = .004$, $\eta_p^2 = .04$. Participants in the deep conditions expected to care more about their responses to the questions than their partner would, $F(1, 99) = 18.05$, $p < .001$, $\eta_p^2 = .15$, but participants in the control conditions did not, $F(1, 101) = 0.07$, $p = .792$, $\eta_p^2 = .0007$.

A 2 (conversation: control, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVA on the partner care measure with repeated measures on the third factor produced a significant main effect of measurement phase, $F(1, 200) = 112.07$, $p < .001$, $\eta_p^2 = .36$, such that participants underestimated partner care. The conversation \times measurement phase interaction effect was non-significant, $F(1, 200) = 0.19$, $p = .667$, $\eta_p^2 = .0009$, indicating that participants underestimated the recipients' care similarly in the deep and

control conditions. Consistent with our predictions, we also observed a significant relationship \times measurement phase interaction effect, $F(1, 200) = 59.11, p < .001, \eta_p^2 = .23$. Although participants who spoke with close others underestimated how much their partner would care about their responses, $F(1, 100) = 4.26, p = .042, \eta_p^2 = .04$, those who spoke with distant strangers did so significantly more, $F(1, 100) = 0.96, p = .329, \eta_p^2 = .01$.

Awkwardness. A 2 (conversation: shallow, deep) \times 2 (relationship: distant, close) \times 2 (measurement phase: expectations, experiences) ANOVA, with repeated measures on the second factor, produced a main effect of measurement phase, $F(1, 200) = 94.19, p < .001, \eta_p^2 = .32$, such that participants overestimated awkwardness, a conversation \times measurement phase interaction effect, $F(1, 200) = 5.60, p = .019, \eta_p^2 = .03$, indicating that deep pairs overestimated awkwardness more than control pairs, and a relationship \times measurement phase interaction effect, $F(1, 200) = 11.08, p = .001, \eta_p^2 = .05$, indicating that distant pairs overestimated awkwardness more than close pairs. Distant pairs predicted that deep conversations would be marginally more awkward than control conversations, $t(100) = -1.84, p = .069, 95\% \text{ CI}_{\text{difference}} = [-1.41, 0.05], d = -0.36$, but experienced no differences in reality, $t(100) = 0.45, p = .655, 95\% \text{ CI}_{\text{difference}} = [-0.46, 0.73], d = 0.09$. Close pairs predicted that deep conversations would be more awkward than control conversations, $t(100) = -2.68, p = .009, 95\% \text{ CI}_{\text{difference}} = [-1.69, -0.25], d = -0.53$, but experienced only marginally greater awkwardness after deep conversations, $t(100) = -1.91, p = .059, 95\% \text{ CI}_{\text{difference}} = [-1.40, 0.03], d = -0.38$.

Enjoyment. For enjoyment, we observed a main effect of measurement phase, $F(1, 196) = 135.31, p < .001, \eta_p^2 = .41$, such that participants underestimated enjoyment, and a relationship \times measurement phase interaction effect, $F(1, 196) = 29.58, p < .001, \eta_p^2 = .13$, such that distant pairs underestimated enjoyment more than close pairs. Distant pairs did not predict differences in

enjoyment between control and deep conversations, $t(100) = -0.48, p = .633, 95\% \text{ CI}_{\text{difference}} = [-0.75, 0.46], d = -0.09$, nor did they experience differences after speaking, $t(100) = -1.09, p = .278, 95\% \text{ CI}_{\text{difference}} = [-0.93, 0.27], d = -0.22$. Close pairs likewise did not predict differences in enjoyment between control and deep conversations, $t(100) = -0.65, p = .518, 95\% \text{ CI}_{\text{difference}} = [-0.76, 0.38], d = -0.13$, nor did they experience differences after speaking, $t(100) = -1.46, p = .147, 95\% \text{ CI}_{\text{difference}} = [-1.06, 0.16], d = -0.29$.

Happiness. For happiness, we observed a main effect of measurement phase, $F(1, 196) = 188.47, p < .001, \eta_p^2 = .49$, such that participants underestimated happiness, and a relationship \times measurement phase interaction effect, $F(1, 196) = 29.96, p < .001, \eta_p^2 = .13$, such that distant pairs underestimated happiness more than close pairs. Distant pairs did not predict differences in happiness between control and deep conversations, $t(100) = -1.21, p = .231, 95\% \text{ CI}_{\text{difference}} = [-0.81, 0.20], d = -0.24$, nor did they experience differences after speaking, $t(100) = -1.24, p = .218, 95\% \text{ CI}_{\text{difference}} = [-0.91, 0.21], d = -0.25$. Close pairs likewise did not predict differences in happiness between control and deep conversations, $t(100) = -0.51, p = .614, 95\% \text{ CI}_{\text{difference}} = [-0.63, 0.37], d = -0.10$, nor did they experience differences after speaking, $t(100) = -1.39, p = .166, 95\% \text{ CI}_{\text{difference}} = [-0.93, 0.16], d = -0.28$.

Mediational analyses. Next we tested whether underestimation of partner care mediated overestimation of awkwardness or underestimation of enjoyment or happiness. Among distant control pairs, the indirect effects were non-significant for awkwardness ($b = 0.24, 95\% \text{ CI} = [-0.30, 0.77]$), non-significant for enjoyment ($b = -0.32, 95\% \text{ CI} = [-0.78, 0.20]$), and non-significant for happiness ($b = -0.31, 95\% \text{ CI} = [-0.64, 0.06]$). Among close control pairs, the indirect effects were non-significant for awkwardness ($b = 0.02, 95\% \text{ CI} = [-0.03, 0.21]$), non-significant for enjoyment ($b = -0.07, 95\% \text{ CI} = [-0.41, 0.08]$), and non-significant for happiness

($b = -0.07$, 95% CI = [-0.30, 0.10]). Among distant deep pairs, the indirect effects were significant for awkwardness ($b = 0.88$, 95% CI = [0.36, 1.44]), non-significant for enjoyment ($b = -0.28$, 95% CI = [-1.13, 0.51]), and significant for happiness ($b = -0.46$, 95% CI = [-0.98, -0.12]). Among close deep pairs, the indirect effects were non-significant for awkwardness ($b = 0.004$, 95% CI = [-0.25, 0.19]), significant for enjoyment ($b = -0.20$, 95% CI = [-0.48, -0.03]), and significant for happiness ($b = -0.14$, 95% CI = [-0.34, -0.02]).

Next we tested whether underestimation of partner care explained differences between close and distant pairs in calibration on the primary measures. Among control pairs, the indirect effects were non-significant for awkwardness ($b = 0.27$, 95% CI = [-0.05, 0.66]), significant for enjoyment ($b = -0.58$, 95% CI = [-0.96, -0.16]), and significant for happiness ($b = -0.55$, 95% CI = [-0.90, -0.25]). Among deep pairs, the indirect effects were also non-significant for awkwardness ($b = 0.34$, 95% CI = [-0.02, 0.79]), significant for enjoyment ($b = -0.41$, 95% CI = [-0.91, -0.01]), and significant for happiness ($b = -0.37$, 95% CI = [-0.71, -0.13]).

Gender Effects

No main effects of measurement phase (prediction vs. experience), or higher-order interactions with measurement phase, were qualified by significant interaction effects with the gender combination of the pair (same-gender versus mixed-gender), $F_s(1, 192) \leq 3.09$, $p_s \geq .080$, η_p^2 s $\leq .02$. The one exception was that mixed-gender pairs were significantly more likely than same-gender pairs to underestimate their happiness, $F(1, 192) = 6.69$, $p = .010$, $\eta_p^2 = .03$.

Experiment 5 Discussion

Method

Participants. Participants were recruited from Amazon's Mechanical Turk ($N = 109$ individuals after exclusions; $M_{\text{age}} = 36.32$; $SD_{\text{age}} = 10.36$; 56.88% female; 70.64% Caucasian) to complete the study in exchange for \$2.00. An additional 3 participants were excluded for failing an attention check.

Procedure. Participants imagined visiting a "social interaction" research lab where they would speak with another study participant. They were told that this discussion would be called the "sharing game."

Then participants were randomly assigned to either the "friend" or "stranger" condition. Participants in the "friend" condition provided the first and last initials of a friend who they know very well and they were told that they would imagine having a conversation with this friend. Participants in the "stranger" condition were told that they would imagine having a conversation with a stranger who they had never met before.

Then participants in both conditions read the set of 20 pre-tested discussion questions (see Appendix), presented in randomized order. Participants selected the 5 questions that they would most like to ask and answer with the other person. Then participants viewed each of the 20 questions separately and reported how interested they would be in asking and answering the question with the other person ($0 = \text{not at all}$; $10 = \text{extremely}$), how much they would care about their own response to the question ($0 = \text{not at all}$; $10 = \text{quite a bit}$), how much they believed the other person would care about the participant's own response to the question ($0 = \text{not at all}$; $10 = \text{quite a bit}$), how awkward they would feel while answering and discussing the question with the

other person (0 = *not at all awkward*; 10 = *very awkward*), and how much they would enjoy answering and discussing the question with the other person (0 = *not at all*; 10 = *very much*).

Participants then completed an attention check item in which they reported whether they imagined speaking with a stranger or a close friend. Participants in the friend condition were also asked to report whether the person whose initials they had entered was somebody real who they considered themselves to be friends with.

Finally, participants reported demographic information and received payment.

Results

Participants in the “friend” condition ($M = 3.10$, $SD = 1.51$) selected more deep questions than participants in the “stranger” condition ($M = 1.79$, $SD = 1.32$), $t(107) = 4.82$, $p < .001$, 95% CI = [0.77, 1.84], $d = 0.92$. Participants in the friend condition selected mostly deep questions, $t(51) = 2.84$, $p = .006$, 95% CI = [2.68, 3.52], $d = 0.39$, whereas participants in the stranger condition selected mostly shallow questions, $t(56) = -4.07$, $p < .001$, 95% CI = [1.44, 2.14], $d = -0.54$. Participants in the “friend” condition also selected questions higher in average intimacy ($M = 5.04$, $SD = 1.24$) based on the pre-test than did participants in the “stranger” condition, ($M = 3.97$, $SD = 1.10$), $t(107) = 4.76$, $p < .001$, 95% CI_{difference} = [0.62, 1.51], $d = 0.91$.

Differences in choice were largely driven by participants’ interest in discussing deep questions. We computed each participant’s mean interest in discussing the 10 shallow questions and the 10 deep questions, separately, and performed a 2 (partner: friend, stranger) \times 2 (conversation: shallow, deep) ANOVA with repeated measures on the second factor and mean interest in discussing shallow or deep questions as the dependent measure. We observed a marginally significant main effect of partner, $F(1, 107) = 3.02$, $p = .085$, $\eta_p^2 = .03$, such that participants expressed greater interest in speaking with friends than strangers, and a significant

main effect of conversation, $F(1, 107) = 27.64, p < .001, \eta_p^2 = .21$, such that participants were more interested in discussing deep questions than shallow questions. Importantly, these main effects were qualified by a significant partner \times conversation interaction effect, $F(1, 107) = 19.10, p < .001, \eta_p^2 = .15$, such that participants in the friend condition were more interested in discussing deep (vs. shallow) questions than participants in the stranger condition (see Figure S1). Participants in the friend and stranger conditions did not differ in their interest in discussing shallow questions, $t(107) = -1.36, p = .178, 95\% \text{ CI}_{\text{difference}} = [-1.31, 0.24], d = -0.26$, but participants in the friend condition were more interested in discussing deep questions, $t(107) = 4.28, p < .001, 95\% \text{ CI}_{\text{difference}} = [0.83, 2.26], d = 0.82$.

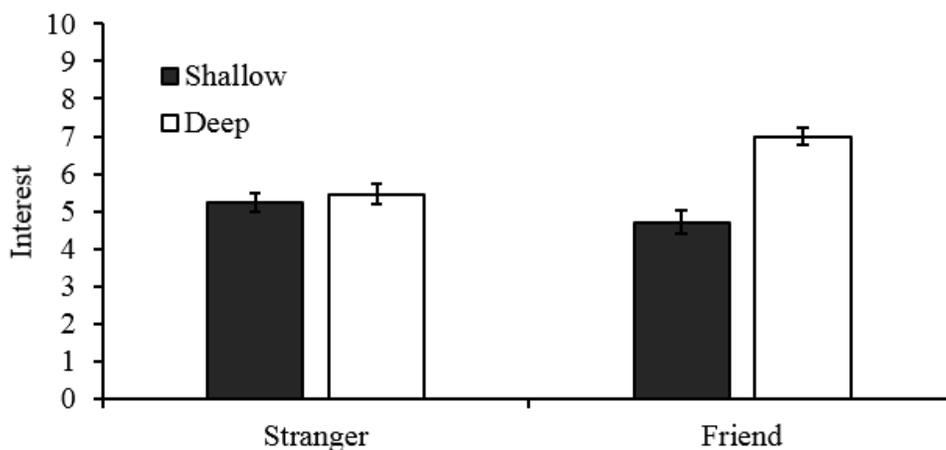


Figure S1. Mean interest across partner (friend or stranger) and conversation (shallow or deep) in the follow-up to Experiment 5. Error bars $\pm 1 SE$.

We observed similar patterns for the participant care, partner care, awkwardness, and enjoyment measures. For each measure, we observed significant main effects of conversation: Participants predicted greater participant care, partner care, awkwardness, and enjoyment for deep questions than for shallow questions, $F_s(1, 107) \geq 20.09, p_s < .001, \eta_p^2_s \geq .16$. Furthermore,

we observed significant partner \times conversation interaction effects for each measure, $F_s(1, 107) \geq 7.52, p_s \leq .007, \eta_p^2_s \geq .07$, indicating that participants in the friend condition expected deep (vs. shallow) questions to produce greater participant care, partner care, and enjoyment, but lower awkwardness, than those in the stranger condition.

Experiment 6a

Full Output from ANOVAs in Main Text

For the partner care measure, we found a significant effect of intimacy, $b = 1.22$, $SE = 0.20$, $t(298.19) = 5.98$, $p < .001$, 95% CI = [0.82, 1.61], indicating that expected care was higher for deep conversations, a significant effect of target, $b = 1.51$, $SE = 0.14$, $t(298.19) = 10.49$, $p < .001$, 95% CI = [1.23, 1.79], indicating that participants thought they would care more about their own responses than their partner would, and a significant intimacy \times target interaction, $b = 0.97$, $SE = 0.29$, $t(298.19) = 3.36$, $p < .001$, 95% CI = [0.40, 1.53], indicating that this self/other caring gap was significantly more pronounced for deep conversations.

For awkwardness, we found an effect of intimacy, $b = -1.19$, $SE = 0.19$, $t(302.11) = -6.25$, $p < .001$, 95% CI = [-1.56, -0.82], indicating that participants reported greater awkwardness for deep conversations, an effect of measurement phase, $b = -1.87$, $SE = 0.27$, $t(302.11) = -6.94$, $p < .001$, 95% CI = [-2.39, -1.34], indicating that participants overestimated awkwardness across both conversations, and an intimacy \times measurement phase interaction, $b = 0.94$, $SE = 0.38$, $t(302.11) = 2.48$, $p = .014$, 95% CI = [0.20, 1.69], indicating that participants in the deep condition overestimated awkwardness more than did participants in the control condition.

For connectedness, we found an effect of intimacy, $b = -1.06$, $SE = 0.13$, $t(292.99) = -8.21$, $p < .001$, 95% CI = [-1.32, -0.81], indicating that participants reported stronger connections for deep conversations, and an effect of measurement phase, $b = 1.02$, $SE = 0.18$, $t(292.99) = 5.58$, $p < .001$, 95% CI = [0.66, 1.38], indicating that participants underestimated connectedness across both conversations. The intimacy \times measurement phase interaction was non-significant, $b = -0.16$, $SE = 0.26$, $t(292.99) = -0.61$, $p = .541$, 95% CI = [-0.66, 0.35].

For happiness, we found an effect of intimacy, $b = -0.24$, $SE = 0.12$, $t(296.00) = -2.06$, $p = .040$, 95% CI = [-0.47, -0.01], indicating that participants reported more happiness for deep conversations, and an effect of measurement phase, $b = 1.17$, $SE = 0.17$, $t(296.00) = 7.03$, $p < .001$, 95% CI = [0.84, 1.49], indicating that participants underestimated happiness across both conversations. These main effects were qualified by a significant intimacy \times measurement phase interaction effect, $b = -0.62$, $SE = 0.23$, $t(296.00) = -2.63$, $p = .009$, 95% CI = [-1.08, -0.16], indicating that participants underestimated their happiness significantly more for the deep conversation than the shallow conversation.

Mediational Analyses at the Level of the Individual

We conducted mediational analyses at the level of the individual, with measurement phase (expectations vs. experiences) as the independent variable, partner care as the mediating variable, and the primary measures as dependent variables, separately for the shallow and deep conversations. For the shallow conversations, partner care significantly mediated the effect of measurement phase (expectations vs. experiences) on connectedness, $b = -0.73$, 95% CI = [-1.06, -0.47], and happiness, $b = -0.30$, 95% CI = [-0.53, -0.11], but not awkwardness, $b = 0.05$, 95% CI = [-0.26, 0.35]. For the deep conversations, partner care significantly mediated the effect of measurement phase on connectedness, $b = -0.66$, 95% CI = [-0.99, -0.39], and happiness, $b = -0.22$, 95% CI = [-0.46, -0.004], but not awkwardness, $b = 0.23$, 95% CI = [-0.32, 0.75].

Analyses Without Exclusions

We reanalyzed the data with all 32 groups included.

Care measures. Replicating the main-text analyses, we found a significant effect of intimacy, $b = 1.16$, $SE = 0.20$, $t(318.13) = 5.81$, $p < .001$, 95% CI = [0.77, 1.55], a significant effect of target, $b = 1.55$, $SE = 0.14$, $t(318.13) = 11.01$, $p < .001$, 95% CI = [1.28, 1.83], and a

significant intimacy \times target interaction, $b = 1.04$, $SE = 0.28$, $t(318.13) = 3.69$, $p < .001$, 95% CI = [0.49, 1.59]. Participants expected to care significantly more about their responses during shallow conversations ($M = 5.05$, $SD = 2.62$) than their partner would ($M = 4.02$, $SD = 2.38$), $b = 1.03$, $SE = 0.20$, $t(318.13) = 5.18$, $p < .001$, 95% CI = [0.64, 1.42], but this gap was significantly larger for deep conversations ($M_s = 7.24$ vs. 5.17, respectively; $SD_s = 2.03$ vs. 2.25), $b = 2.07$, $SE = 0.20$, $t(318.13) = 10.40$, $p < .001$, 95% CI = [1.68, 2.46].

Awkwardness. We found an effect of intimacy, $b = -1.20$, $SE = 0.18$, $t(320.59) = -6.60$, $p < .001$, 95% CI = [-1.55, -0.84], an effect of measurement phase, $b = -1.84$, $SE = 0.26$, $t(320.59) = -7.15$, $p < .001$, 95% CI = [-2.34, -1.33], and an intimacy \times measurement phase interaction, $b = 0.99$, $SE = 0.36$, $t(320.59) = 2.73$, $p = .007$, 95% CI = [0.28, 1.70]. Although participants overestimated how awkward their shallow conversations would be, $b = -0.84$, $SE = 0.26$, $t(320.59) = -3.29$, $p = .001$, 95% CI = [-1.35, -0.34], they did so significantly more for deep conversations, $b = -1.84$, $SE = 0.26$, $t(320.59) = -7.15$, $p < .001$, 95% CI = [-2.34, -1.33].

Connectedness. We found an effect of intimacy, $b = -1.07$, $SE = 0.13$, $t(311.92) = -8.56$, $p < .001$, 95% CI = [-1.32, -0.83], and an effect of measurement phase, $b = 1.07$, $SE = 0.18$, $t(311.92) = 6.03$, $p < .001$, 95% CI = [0.72, 1.42]. The intimacy \times measurement phase interaction was non-significant, $b = -0.16$, $SE = 0.25$, $t(311.92) = -0.62$, $p = .534$, 95% CI = [-0.65, 0.33]. Participants expected to feel more connected to their deep conversation partner, $b = 1.00$, $SE = 0.18$, $t(311.92) = 5.61$, $p < .001$, 95% CI = [0.65, 1.34], and felt significantly more connected to this person after speaking, $b = 1.15$, $SE = 0.18$, $t(311.92) = 6.49$, $p < .001$, 95% CI = [0.80, 1.50].

Participants reported significantly stronger connections to their deep conversation partner than their shallow conversation partner ($M = 1.55$, $SD = 3.04$), one-sample $t(127) = 5.76$, $p < .001$, 95% CI = [1.02, 2.08], $d = 0.51$.

Happiness. We found a marginally significant effect of intimacy, $b = -0.22$, $SE = 0.11$, $t(315.13) = -1.90$, $p = .058$, 95% CI = [-0.44, 0.01], and a significant effect of measurement phase, $b = 1.05$, $SE = 0.16$, $t(315.13) = 6.49$, $p < .001$, 95% CI = [0.74, 1.37]. The intimacy \times measurement phase interaction was significant, $b = -0.47$, $SE = 0.23$, $t(315.13) = -2.04$, $p = .042$, 95% CI = [-0.02, -0.02], indicating that participants underestimated their happiness significantly more for the deep conversation than the shallow conversation. Participants expected to feel equally happy after their shallow and deep conversations, $b = -0.02$, $SE = 0.16$, $t(315.13) = -0.10$, $p = .923$, 95% CI = [-0.33, 0.30], yet felt significantly happier after their deep conversations, $b = 0.45$, $SE = 0.16$, $t(315.13) = 2.79$, $p = .006$, 95% CI = [0.14, 0.77].

Mediational analyses. For the shallow conversations, partner care significantly mediated the effect of measurement phase (expectations vs. experiences) on connectedness, $b = -0.68$, $SE = 0.22$, 95% CI = [-1.17, -0.31], and happiness, $b = -0.35$, $SE = 0.18$, 95% CI = [-0.72, -0.02], but not awkwardness, $b = 0.07$, $SE = 0.24$, 95% CI = [-0.42, 0.51]. For the deep conversations, partner care significantly mediated the effect of measurement phase on connectedness, $b = -0.97$, $SE = 0.28$, 95% CI = [-1.54, -0.45], but not awkwardness, $b = 0.67$, $SE = 0.36$, 95% CI = [-0.01, 1.35], or happiness, $b = -0.28$, $SE = 0.23$, 95% CI = [-0.71, 0.19].

Desired conversation intimacy. Participants reported that their typical conversations with strangers ($M = 3.61$, $SD = 2.37$) were significantly less intimate than they would ideally prefer ($M = 5.46$, $SD = 2.53$), $paired\ t(127) = -9.30$, $p < .001$, 95% CI_{difference} = [-2.25, -1.46], $d = -0.75$.

Gender Effects

Same-gender pairs were significantly more likely than mixed-gender pairs to underestimate their own care, their partner's care, their connections, and their happiness, $bs \geq 0.61$, $ts \geq 2.24$, $ps \leq .026$, but not the awkwardness of their conversations, $b = 0.09$, $SE = 0.40$, $t(295.47) = 0.24$, 95%

CI = [-0.68, 0.86]. Notably, both same-gender pairs ($ts > 5.58, ps < .001$) and mixed-gender pairs ($ts \geq 2.53, ps \leq .013$) significantly underestimated the positivity of their conversations on each of the partner-care, awkwardness, connection, and happiness measures. Differences in miscalibration between shallow and deep conversations did not differ significantly between same-gender and mixed-gender pairs on any measure, $-1.03 \leq bs \leq 0.61, ts \leq 1.60, ps \geq .110$.

Experiments 7a & 7b

Analyses Without Exclusions (Experiment 7a)

We reanalyzed the data from Experiment 7a with all 100 participants included. Participants in the caring condition ($M = 3.00$, $SD = 1.46$) selected a greater number of deep questions than participants in the uncaring condition ($M = 2.02$, $SD = 1.68$), $t(98) = 3.11$, $p = .002$, 95% $CI_{\text{difference}} = [0.36, 1.60]$, $d = 0.62$. Furthermore, participants in the caring condition ($M = 4.90$, $SD = 1.24$) selected questions with greater average intimacy than participants in the uncaring condition ($M = 4.24$, $SD = 1.44$), $t(98) = 2.49$, $p = .015$, 95% $CI_{\text{difference}} = [0.13, 1.20]$, $d = 0.50$.

In a 2 (partner: caring, uncaring) \times 2 (question type: shallow, deep) ANOVA with repeated measures on the second factor and mean desire to discuss shallow or deep questions as the dependent measure, we observed a significant partner \times question type interaction effect, $F(1, 98) = 9.08$, $p = .003$, $\eta_p^2 = .08$: Participants in the caring and uncaring conditions did not differ in their desire to discuss shallow questions, $t(98) = -0.33$, $p = .744$, 95% $CI_{\text{difference}} = [-0.90, 0.65]$, $d = -0.07$, but participants in the caring condition ($M = 6.56$, $SD = 1.91$) expressed greater desire to discuss deep questions than participants in the uncaring condition, $t(98) = 3.49$, $p < .001$, 95% $CI_{\text{difference}} = [0.59, 2.15]$, $d = 0.70$.

Analyses Without Exclusions (Experiment 7b)

We re-analyzed the data including all 160 participants. The number of deep questions selected varied significantly across care conditions, $F(2, 157) = 4.10$, $p = .018$, $\eta_p^2 = .05$. Underestimation participants ($M = 2.37$, $SD = 1.15$) selected more deep questions than participants in the overestimation condition ($M = 1.70$, $SD = 1.27$), $t(157) = -2.74$, $p = .007$, 95% $CI_{\text{difference}} = [-1.15, -0.19]$, $d = -0.53$. Participants in the control condition ($M = 2.21$, $SD = 1.36$)

also selected more deep questions than participants in the overestimation condition, $t(157) = -2.07, p = .040, 95\% \text{ CI}_{\text{difference}} = [-0.99, -0.02], d = -0.40$. Choice of deep questions did not differ significantly between participants in the underestimation condition and participants in the control condition, $t(157) = -0.65, p = .518, 95\% \text{ CI}_{\text{difference}} = [-0.64, 0.33], d = -0.13$.

The average intimacy of the selected questions varied significantly across care conditions, $F(2, 157) = 3.88, p = .023, \eta_p^2 = .05$. Participants in the underestimation condition selected questions with higher average intimacy ($M = 4.50, SD = 1.07$) than participants in the overestimation condition ($M = 3.92, SD = 1.11$), $t(157) = -2.79, p = .006, 95\% \text{ CI}_{\text{difference}} = [-1.00, -0.17], d = -0.54$. However, participants in the control condition ($M = 4.23, SD = 1.11$) did not differ significantly from participants in the underestimation condition, $t(157) = -1.30, p = .194, 95\% \text{ CI}_{\text{difference}} = [-0.70, 0.14], d = -0.25$, or the overestimation condition, $t(157) = 1.46, p = .148, 95\% \text{ CI}_{\text{difference}} = [-0.11, 0.73], d = 0.28$.

We then performed a 2 (care: underestimation, overestimation) \times 2 (question type: shallow, deep) ANOVA with repeated measures on the second factor and mean interest in discussing shallow or deep questions as the dependent measure. We observed a significant main effect of care, $F(1, 106) = 6.37, p = .013, \eta_p^2 = .06$, such that participants in the underestimation condition were more interested in discussing the questions than participants in the overestimation condition, and a significant main effect of question type, $F(1, 106) = 6.47, p = .012, \eta_p^2 = .06$, such that participants were more interested in discussing deep questions than shallow questions. The care \times question type interaction effect was non-significant, $F(1, 106) = 1.55, p = .216, \eta_p^2 = .01$. Participants in the underestimation condition did not differ from those in the overestimation condition in their interest in discussing shallow questions, $t(157) = -1.56, p = .121, 95\%$

CI_{difference} = [-1.47, 0.17], $d = -0.30$, but they were significantly more interested in discussing deep questions, $t(157) = -2.97$, $p = .003$, 95% CI_{difference} = [-1.97, -0.40], $d = -0.57$.

Actor-Partner Interdependence Model Analyses

Throughout our experiments, we performed exploratory analyses using the actor-partner interdependence model (Kenny, Kashy, & Cook, 2006) to test whether differences between participants' expectations and experiences were more strongly associated with the participants' own expectations (as our theory would suggest) or their partners' expectations. We inputted one's own expectations and one's partner's expectations as predictors and one's own miscalibration (experiences – expectations) as the outcome variable, separately for each of the primary measures. We found consistent evidence that miscalibration was associated with one's own expectations: participants who expected their partners to care very little were especially likely to underestimate their partners' care; participants who expected their conversations to be especially awkward were especially likely to overestimate the awkwardness of their conversations; participants who expected to form especially weak connections to their partner were especially likely to underestimate the strength of their connections; and participants who expected to feel especially unhappy with their conversations were especially likely to underestimate their happiness. In contrast, we found little evidence that miscalibration was associated with one's partner's expectations. The few associations that were statistically significant varied in direction across measures and experiments, and in every case were weaker than the associations with one's own expectations (see Table S2).

Table S2. Actor-partner interdependence analyses for the experiments in the main text.

	DV = Care or Interest (Experiences – Predictions)	DV = Awkwardness (Experiences – Predictions)	DV = Connectedness or Enjoyment (Experiences – Predictions)	DV = Happiness (Experiences – Predictions)
Exp. 1a	<i>Actor-to-partner interest:</i> Intercept: $b = 7.00, t = 8.37, p < .001$ Actor: $b = -0.78, t = -10.23, p < .001$ Partner: $b = 0.06, t = 0.69, p = .496$	Intercept: $b = 0.94, t = 1.83, p = .081$ Actor: $b = -0.66, t = -6.90, p < .001$ Partner: $b = -0.07, t = -0.71, p = .485$	Intercept: $b = 3.98, t = 5.04, p < .001$ Actor: $b = -0.56, t = -6.38, p < .001$ Partner: $b = 0.23, t = 2.60, p = .012$	Intercept: $b = 6.79, t = 6.95, p < .001$ Actor: $b = -0.79, t = -8.34, p < .001$ Partner: $b = 0.09, t = 0.98, p = .332$
	<i>Partner-to-actor interest:</i> Intercept: $b = 6.91, t = 6.38, p < .001$ Actor: $b = -0.79, t = -6.63, p < .001$ Partner: $b = 0.02, t = 0.20, p = .846$			
Exp. 1b	<i>Actor-to-partner care:</i> Intercept: $b = 4.93, t = 7.58, p < .001$ Actor: $b = -0.58, t = -7.60, p < .001$ Partner: $b = 0.11, t = 1.64, p = .105$	Intercept: $b = -0.38, t = -0.50, p = .622$ Actor: $b = -0.40, t = -5.04, p < .001$ Partner: $b = 0.02, t = 0.26, p = .796$	Intercept: $b = 4.83, t = 6.82, p < .001$ Actor: $b = -0.62, t = -8.33, p < .001$ Partner: $b = 0.10, t = 1.30, p = .198$	Intercept: $b = 5.48, t = 8.90, p < .001$ Actor: $b = -0.66, t = -9.90, p < .001$ Partner: $b = 0.09, t = 1.37, p = .175$
	<i>Partner-to-actor care:</i> Intercept: $b = 5.80, t = 7.59, p < .001$ Actor: $b = -0.71, t = -9.75, p < .001$ Partner: $b = 0.08, t = 1.00, p = .320$			
Exp. 1c	—	Intercept: $b = 0.26, t = 2.33, p = .024$ Actor: $b = -0.78, t = 11.76, p < .001$ Partner: $b = 0.05, t = 0.67, p = .505$	Intercept: $b = 6.66, t = 9.97, p < .001$ Actor: $b = -0.77, t = -10.45, p < .001$ Partner: $b = -0.03, t = -0.31, p = .757$	—
Exp. 2	—	<i>Shallow:</i> Intercept: $b = 0.27, t = 1.52, p = .135$ Actor: $b = -0.70, t = -11.23, p < .001$ Partner: $b = 0.06, t = 0.91, p = .367$	<i>Shallow:</i> Intercept: $b = 3.50, t = 3.23, p = .002$ Actor: $b = -0.48, t = -3.51, p = .001$ Partner: $b = 0.06, t = 0.46, p = .649$	—
		<i>Deep:</i> Intercept: $b = -0.49, t = -0.71, p = .485$ Actor: $b = -0.55, t = -6.86, p < .001$ Partner: $b = 0.05, t = 0.59, p = .561$	<i>Deep:</i> Intercept: $b = 3.15, t = 4.09, p < .001$ Actor: $b = -0.57, t = -5.54, p < .001$ Partner: $b = 0.24, t = 2.31, p = .024$	
Exp. 3	—	<i>Control:</i> Intercept: $b = 0.42, t = 1.07, p = .289$ Actor: $b = -0.63, t = -8.14, p < .001$ Partner: $b = 0.13, t = 1.65, p = .102$	<i>Control:</i> Intercept: $b = 1.25, t = 1.93, p = .059$ Actor: $b = -0.22, t = -2.50, p = .014$ Partner: $b = 0.17, t = 1.91, p = .059$	<i>Control:</i> Intercept: $b = 1.97, t = 2.52, p = .015$ Actor: $b = -0.32, t = -3.55, p = .001$ Partner: $b = 0.13, t = 1.37, p = .175$
		<i>Deep:</i> Intercept: $b = 0.44, t = 0.93, p = .356$ Actor: $b = -0.56, t = -7.48, p < .001$ Partner: $b = -0.01, t = -0.18, p = .860$	<i>Deep:</i> Intercept: $b = 2.77, t = 2.97, p = .005$ Actor: $b = -0.34, t = -3.50, p = .001$ Partner: $b = 0.03, t = 0.27, p = .785$	<i>Deep:</i> Intercept: $b = 3.39, t = 3.71, p = .001$ Actor: $b = -0.30, t = -3.37, p = .001$ Partner: $b = -0.06, t = -0.65, p = .516$
Exp. 4a	<i>Shallow, self care:</i> Intercept: $b = 4.35, t = 4.26, p < .001$ Actor: $b = -0.74, t = -7.25, p < .001$ Partner: $b = 0.01, t = 0.10, p = .918$			
	<i>Shallow, other care:</i> Intercept: $b = 6.58, t = 9.65, p < .001$ Actor: $b = -0.82, t = -9.83, p < .001$ Partner: $b = -0.11, t = -1.29, p = .199$	<i>Shallow:</i> Intercept: $b = 1.02, t = 3.25, p = .002$ Actor: $b = -0.85, t = -10.93, p < .001$ Partner: $b = -0.02, t = -0.25, p = .801$	<i>Shallow:</i> Intercept: $b = 6.28, t = 12.00, p < .001$ Actor: $b = -0.60, t = -9.00, p < .001$ Partner: $b = -0.17, t = -2.61, p = .010$	<i>Shallow:</i> Intercept: $b = 6.21, t = 10.88, p < .001$ Actor: $b = -0.61, t = -9.42, p < .001$ Partner: $b = -0.99, t = -1.53, p = .129$
	<i>Deep, self care:</i> Intercept: $b = 4.63, t = 2.92, p = .005$ Actor: $b = -0.48, t = -3.61, p < .001$ Partner: $b = -0.18, t = -1.35, p = .182$	<i>Deep:</i> Intercept: $b = 0.61, t = 0.94, p = .350$ Actor: $b = -0.54, t = -6.43, p < .001$ Partner: $b = 0.01, t = 0.08, p = .939$	<i>Deep:</i> Intercept: $b = 6.27, t = 7.72, p < .001$ Actor: $b = -0.65, t = -8.06, p < .001$ Partner: $b = -0.99, t = -1.23, p = .221$	<i>Deep:</i> Intercept: $b = 5.19, t = 6.08, p < .001$ Actor: $b = -0.55, t = -6.31, p < .001$ Partner: $b = -0.05, t = -0.61, p = .546$
	<i>Deep, other care:</i> Intercept: $b = 6.42, t = 7.85, p < .001$ Actor: $b = -0.72, t = -7.59, p < .001$ Partner: $b = -0.06, t = -0.64, p = .526$			

Table S2 continued.

	DV = Care or Interest (Experiences – Predictions)	DV = Awkwardness (Experiences – Predictions)	DV = Connectedness or Enjoyment (Experiences – Predictions)	DV = Happiness (Experiences – Predictions)
Exp. 4b	<i>Shallow, self care:</i> Intercept: $b = 4.06, t = 3.96, p < .001$ Actor: $b = -0.54, t = -5.13, p < .001$ Partner: $b = 0.04, t = 0.37, p = .711$			
	<i>Shallow, other care:</i> Intercept: $b = 4.75, t = 6.84, p < .001$ Actor: $b = -0.58, t = -7.20, p < .001$ Partner: $b = 0.05, t = 0.62, p = .538$	<i>Shallow:</i> Intercept: $b = 0.81, t = 2.47, p = .017$ Actor: $b = -0.62, t = -9.51, p < .001$ Partner: $b = -0.14, t = -2.15, p = .034$	<i>Shallow:</i> Intercept: $b = 5.81, t = 7.24, p < .001$ Actor: $b = -0.66, t = -7.31, p < .001$ Partner: $b = -0.01, t = -0.16, p = .877$	<i>Shallow:</i> Intercept: $b = 5.00, t = 8.30, p < .001$ Actor: $b = -0.55, t = -8.86, p < .001$ Partner: $b = 0.005, t = 0.08, p = .940$
	<i>Deep, self care:</i> Intercept: $b = 5.59, t = 4.13, p < .001$ Actor: $b = -0.69, t = -5.83, p < .001$ Partner: $b = -0.09, t = -0.76, p = .4448$	<i>Deep:</i> Intercept: $b = 0.95, t = 1.14, p = .259$ Actor: $b = -0.72, t = -8.27, p < .001$ Partner: $b = 0.03, t = 0.36, p = .719$	<i>Deep:</i> Intercept: $b = 4.27, t = 3.78, p < .001$ Actor: $b = -0.65, t = -6.63, p < .001$ Partner: $b = 0.14, t = 1.46, p = .149$	<i>Deep:</i> Intercept: $b = 4.11, t = 4.39, p < .001$ Actor: $b = -0.43, t = -5.29, p < .001$ Partner: $b = -0.04, t = -0.43, p = .669$
	<i>Deep, other care:</i> Intercept: $b = 5.47, t = 6.92, p < .001$ Actor: $b = -0.70, t = -8.06, p < .001$ Partner: $b = 0.04, t = 0.46, p = .645$			
	<i>Control, unacquainted, self care:</i> Intercept: $b = 3.36, t = 5.05, p < .001$ Actor: $b = -0.43, t = -6.20, p < .001$ Partner: $b = -0.02, t = -0.30, p = .763$			
	<i>Control, unacquainted, other care:</i> Intercept: $b = 4.39, t = 5.87, p < .001$ Actor: $b = -0.50, t = -5.65, p < .001$ Partner: $b = 0.02, t = 0.25, p = .807$			
	<i>Control, acquainted, self care:</i> Intercept: $b = 1.51, t = 1.66, p = .104$ Actor: $b = -0.40, t = -4.68, p < .001$ Partner: $b = 0.17, t = 1.96, p = .053$	<i>Control, unacquainted:</i> Intercept: $b = 0.44, t = 1.37, p = .178$ Actor: $b = -0.63, t = -8.90, p < .001$ Partner: $b = 0.05, t = 0.71, p = .480$	<i>Control, unacquainted:</i> Intercept: $b = 4.38, t = 5.73, p < .001$ Actor: $b = -0.50, t = -6.31, p < .001$ Partner: $b = 0.05, t = 0.66, p = .508$	<i>Control, unacquainted:</i> Intercept: $b = 3.21, t = 4.48, p < .001$ Actor: $b = -0.43, t = -5.35, p < .001$ Partner: $b = 0.17, t = 2.13, p = .036$
	<i>Control, acquainted, other care:</i> Intercept: $b = 3.54, t = 3.43, p = .001$ Actor: $b = -0.45, t = -5.36, p < .001$ Partner: $b = -0.01, t = -0.08, p = .937$	<i>Control, acquainted:</i> Intercept: $b = 0.40, t = 1.50, p = .139$ Actor: $b = -0.55, t = -7.93, p < .001$ Partner: $b = -0.01, t = -0.13, p = .895$	<i>Control, acquainted:</i> Intercept: $b = 2.04, t = 2.20, p = .033$ Actor: $b = -0.33, t = -4.37, p < .001$ Partner: $b = 0.12, t = 1.65, p = .103$	<i>Control, acquainted:</i> Intercept: $b = 2.37, t = 2.26, p = .029$ Actor: $b = -0.31, t = -3.33, p = .001$ Partner: $b = 0.07, t = 0.74, p = .464$
	<i>Deep, unacquainted, self care:</i> Intercept: $b = 1.44, t = 1.74, p = .089$ Actor: $b = -0.32, t = 3.66, p < .001$ Partner: $b = 0.13, t = 1.44, p = .152$	<i>Deep, unacquainted:</i> Intercept: $b = -0.10, t = -0.28, p = .785$ Actor: $b = -0.63, t = -10.46, p < .001$ Partner: $b = 0.10, t = 1.61, p = .111$	<i>Deep, unacquainted:</i> Intercept: $b = 4.55, t = 5.75, p < .001$ Actor: $b = -0.57, t = -7.91, p < .001$ Partner: $b = 0.14, t = 1.96, p = .054$	<i>Deep, unacquainted:</i> Intercept: $b = 4.30, t = 4.54, p < .001$ Actor: $b = -0.53, t = -6.40, p < .001$ Partner: $b = 0.12, t = 1.47, p = .145$
	<i>Deep, unacquainted, other care:</i> Intercept: $b = 4.87, t = 6.33, p < .001$ Actor: $b = -0.60, t = -6.71, p < .001$ Partner: $b = 0.06, t = 0.69, p = .489$	<i>Deep, acquainted:</i> Intercept: $b = -0.001, t = -0.002, p = .999$ Actor: $b = -0.45, t = -5.85, p < .001$ Partner: $b = 0.13, t = 1.69, p = .094$	<i>Deep, acquainted:</i> Intercept: $b = 4.36, t = 4.52, p < .001$ Actor: $b = -0.48, t = -6.01, p < .001$ Partner: $b = 0.02, t = 0.23, p = .820$	<i>Deep, acquainted:</i> Intercept: $b = 4.57, t = 5.28, p < .001$ Actor: $b = -0.40, t = -5.47, p < .001$ Partner: $b = -0.86, t = -1.17, p = .245$
	<i>Deep, acquainted, self care:</i> Intercept: $b = 4.21, t = 5.39, p < .001$ Actor: $b = -0.42, t = -4.57, p < .001$ Partner: $b = -0.08, t = -0.85, p = .397$			
	<i>Deep, acquainted, other care:</i> Intercept: $b = 4.78, t = 5.59, p < .001$ Actor: $b = -0.51, t = -6.10, p < .001$ Partner: $b = -0.03, t = -0.30, p = .764$			

Exp.
5

Table S2 continued.

	DV = Care or Interest (Experiences – Predictions)	DV = Awkwardness (Experiences – Predictions)	DV = Connectedness or Enjoyment (Experiences – Predictions)	DV = Happiness (Experiences – Predictions)
Exp. 6a	<i>Shallow, self care:</i> Intercept: $b = 2.41, t = 3.60, p = .001$ Actor: $b = -0.37, t = -4.51, p < .001$ Partner: $b = -0.004, t = -0.04, p = .966$			
	<i>Shallow, other care:</i> Intercept: $b = 3.36, t = 6.18, p < .001$ Actor: $b = -0.46, t = -5.46, p < .001$ Partner: $b = 0.01, t = 0.16, p = .876$	<i>Shallow:</i> Intercept: $b = 1.26, t = 3.06, p = .005$ Actor: $b = -0.37, t = -5.12, p < .001$ Partner: $b = -0.17, t = -2.32, p = .023$	<i>Shallow:</i> Intercept: $b = 3.24, t = 5.51, p < .001$ Actor: $b = -0.50, t = -5.40, p < .001$ Partner: $b = -0.02, t = -0.18, p = .857$	<i>Shallow:</i> Intercept: $b = 2.85, t = 4.31, p < .001$ Actor: $b = -0.32, t = -3.72, p < .001$ Partner: $b = -0.09, t = -1.09, p = .276$
	<i>Deep, self care:</i> Intercept: $b = 3.00, t = 3.45, p = .001$ Actor: $b = -0.41, t = -5.15, p < .001$ Partner: $b = 0.02, t = 0.21, p = .835$	<i>Deep:</i> Intercept: $b = 1.08, t = 1.49, p = .141$ Actor: $b = -0.55, t = -7.25, p < .001$ Partner: $b = 0.04, t = 0.51, p = .608$	<i>Deep:</i> Intercept: $b = 4.43, t = 6.23, p < .001$ Actor: $b = -0.51, t = -6.44, p < .001$ Partner: $b = -0.10, t = -1.28, p = .203$	<i>Deep:</i> Intercept: $b = 3.79, t = 5.19, p < .001$ Actor: $b = -0.40, t = -5.08, p < .001$ Partner: $b = -0.08, t = -1.07, p = .288$
	<i>Deep, other care:</i> Intercept: $b = 5.67, t = 8.35, p < .001$ Actor: $b = -0.60, t = -7.54, p < .001$ Partner: $b = -0.11, t = -1.37, p = .173$			