

Rationales for Indirect Speech: The Theory of the Strategic Speaker

James J. Lee and Steven Pinker
Harvard University

Speakers often do not state requests directly but employ innuendos such as *Would you like to see my etchings?* Though such indirectness seems puzzlingly inefficient, it can be explained by a theory of the *strategic speaker*, who seeks plausible deniability when he or she is uncertain of whether the hearer is cooperative or antagonistic. A paradigm case is bribing a policeman who may be corrupt or honest: A veiled bribe may be accepted by the former and ignored by the latter. Everyday social interactions can have a similar payoff structure (with emotional rather than legal penalties) whenever a request is implicitly forbidden by the relational model holding between speaker and hearer (e.g., bribing an honest maitre d', where the reciprocity of the bribe clashes with his authority). Even when a hearer's willingness is known, indirect speech offers higher-order plausible deniability by preempting certainty, gossip, and common knowledge of the request. In supporting experiments, participants judged the intentions and reactions of characters in scenarios that involved fraught requests varying in politeness and directness.

Keywords: pragmatics, politeness, relational models, common knowledge, indirect speech, psycholinguistics

In the film *Schindler's List* (Spielberg, 1993), after some of Schindler's Jewish workers have been deported to Auschwitz, he negotiates with an SS commandant for their release. The commandant says, "It is not my task to interfere with the processes that take place down here. What makes you think that I can help you?" Schindler replies, "Allow me to express the reason," and empties a satchel of diamonds onto the table. The commandant says, "I could have you arrested." Schindler's response: "I'm protected by powerful friends. You should know that." This leaves the commandant at a loss. Finally he says, "I'm not saying that I am accepting them. All I say is that I am not comfortable with them on the table." The commandant then scoops up the diamonds and jams them into his pocket.

No one who watches this conversation can avoid filling in the subtext of bribes, threats, solicitations, and denials that pass between these characters. Yet the terms of those transactions are never stated explicitly. Schindler does not say, "If you release my workers, I will give you the diamonds" or "If you have me arrested, I will have you punished." Nor does the commandant say, "If you reward me, I will release the workers" or "I accept your offer but will deny it to third parties."

Linguists refer to such innuendoes as *off-record indirect speech acts* (P. Brown & Levinson, 1987). They may be distinguished from "on-record" indirect speech acts such as *Can you pass the salt?*, which have become so conventionalized that hearers rarely

perceive the literal meaning (Gibbs, 1983; Holtgraves, 1994). Here are some other recognizable examples of off-record indirect speech:

- "I hear you're the foreman of the jury in the Soprano trial. It's an important civic responsibility. You have a wife and kids. We know you'll do the right thing." [a threat]
- "Gee, officer. I was thinking that maybe the best thing would be to take care of the ticket here, without going through a lot of paperwork." [a bribe]
- "Would you like to come up and see my etchings?" [a sexual advance]
- "We're counting on you to show leadership in our Campaign for the Future." [a solicitation for a donation]

Off-record indirect speech is a phenomenon with both scientific and practical importance. The puzzle for social psychology and psycholinguistics is why people so often communicate in ways that seem inefficient and error-prone rather than stating their intentions succinctly, clearly, and unambiguously. The practical importance lies in the many legal cases that hinge on the interpretation of indirect speech. Examples include the 1991 Senate confirmation hearing for Supreme Court nominee Clarence Thomas, which debated whether his sexual banter with supervisee Anita Hill was a sexual come-on; the 2008 arrest of Massachusetts State Senator Dianne Wilkerson, whose acceptance of \$2,000 "in appreciation of her efforts" to obtain a liquor license for a client was treated as a case of bribery; and the 2009 offer by Robert Halderman to sell a screenplay to David Letterman depicting his sexual relationships with staffers, which led to Halderman's arrest for attempted blackmail. The ambiguities of indirect speech are also a major source of misunderstanding and conflict within personal relationships (Tannen, 1991).

Indirect speech has long been studied by linguists and philosophers (Cole & Morgan, 1975; Grice, 1975; Horn, 2003; Lakoff, 1973), who have documented the relationship between the form of

James J. Lee and Steven Pinker, Department of Psychology, Harvard University.

This research was supported by National Institutes of Health Grant HD-18381. We thank Alex Levin and Jennifer Kan for helping to design materials and run participants in Experiment 2.

Correspondence concerning this article should be addressed to James J. Lee or to Steven Pinker, Department of Psychology, Harvard University, William James Hall 970, 33 Kirkland Street, Cambridge, MA 02138. E-mail: jameslee@wjh.harvard.edu or pinker@wjh.harvard.edu

an utterance and the intended meaning of the speaker, and the processes by which speakers encode these meanings and hearers recover them. A widespread assumption is that people tacitly respect Grice's (1975) *cooperative principle*: that a speaker and hearer cooperate to move a conversation forward, the hearer filling in whatever propositions are necessary to preserve the assumption that the speaker is trying to be informative, truthful, clear, and relevant.

This literature says little about the social psychological question of why speakers veil their utterances in the first place, given that indirect speech, by definition, flouts the cooperative principle's maxims to be clear, concise, truthful, and relevant. The social motives behind indirect speech have been explored in important synthetic works by P. Brown and Levinson (1987) and Clark (1996). In their theory of politeness, P. Brown and Levinson extended the cooperative principle to the speaker and hearer's cooperation in maintaining *face* (from the idiom *to save face*)—the degree of approval and autonomy that a person can claim in social interactions (Goffman, 1967). Because a speaker's request for attention or favors is a threat to the hearer's face, speakers soften their requests with several kinds of politeness. These include assurances of sympathy (*positive politeness*, as in compliments and terms of endearment and familiarity) and acknowledgments of deference (*negative politeness*, as in apologies, hedges, and inter-rogatives). In this theory, politeness strategies are arranged on a continuum of face-restoring power: positive politeness, negative politeness, on-record indirect speech, and off-record indirect speech. Clark's *joint action theory* also places cooperation at the center of language use. He introduces the theory as follows:

Language use is really a form of *joint action*. A joint action is one that is carried out by an ensemble of people acting in coordination with each other. As simple examples, think of two people waltzing, paddling a canoe, playing a piano duet, or making love. (Clark, 1996, p. 3)

A key rationale for indirect speech in this theory is the joint striving for *equity*. If costs and benefits are not balanced for a speaker or hearer, it threatens their face, leads to distress, and sets up a shared goal of redressing the imbalance through compensating action or speech or a reinterpretation of the goods at stake (Walster, Walster, & Berscheid, 1978). In one of Clark's examples,

Alan offers Barbara some Manzanilla sherry, and she accepts. When Alan proposes the offer, he puts his face at risk. What if she takes the sherry without adequate recompense? And when Barbara takes up his proposal, she puts her own face at risk. What if she cannot repay him for the favor? (Clark, 1996, p. 294)

Clark noted that Barbara might save face by saying *Yes please* (deferring to his autonomy) or *That'd be lovely* (augmenting his self-worth). In politeness theory, these would be classified as examples of negative politeness and on-record indirect speech, respectively.

Despite the many theoretical and empirical achievements of these pure-cooperation theories, they are less successful in explaining certain aspects of off-record indirect speech. Prima facie, pure-cooperation theories appear inconsistent with the fact that off-record indirect speech is often accompanied by significant interpersonal *conflict*, as in the Thomas, Wilkerson, and Letterman legal cases. Nor

does the complete harmony of joint actions such as waltzing, canoeing, or sex seem like an apt analogy to the substantial tension and misunderstanding that surrounds indirect speech. Just consider the kind of conversation that often *precedes* sex, namely, the nervous, tentative, and wary exchanges that go into seducing and propositioning. Emotionally fraught and potentially conflictual propositions, such as bribes, threats, and sexual come-ons, are rarely discussed in the literature on cooperation theories, despite their being among the prime scenarios in which off-record indirect speech is called for. Such come-ons may be highly indirect, but they are unlikely to be clad in the kinds of constructions that protect the hearer's face in other requests, such as *Please, Do you think you might, or I'm sorry to have to ask but*.

Empirical research designed to test politeness theory also casts doubt on its conjecture that off-record indirect speech lies at the maximally polite end of a continuum of strategies (Dillard, Wilson, Tusing, & Kinney, 1997; Holtgraves & Yang, 1990). Raters often deem off-record indirect speech to be *less* polite than negative (deferential) politeness, and sometimes as downright rude (e.g., as in *Didn't I already ask you to empty the dishwasher?*).

A second empirical complication for pure-cooperation theories is that positive and negative politeness and the face threats that call for them have been found to differ not just in degree but also in kind. Positive politeness (sympathy) tends to accompany threats to solidarity, such as criticizing a friend, whereas negative politeness (deference) tends to accompany threats to power, such as an onerous request. The neglect of qualitative differences among kinds of relationships also raises questions about the appeal to equity in Clark's (1996) theory. If, following Alan's offer of sherry to Barbara at a dinner party, she were to attempt to restore equity by offering to pay him for the drink, or if he were to ask her to return the favor in the form of a sexual kiss, the exchange would result not in a reduction of emotional tension but an elevation, experienced as puzzlement, awkwardness, or shock.

We suggest that the limitations of traditional cooperation theories in explaining off-record indirect speech acts may be remedied by two insights from evolutionary biology. The first is that social relationships among conspecifics never involve a perfect overlap of interests (and hence full cooperation) but always entail at least a partial conflict of interest (Trivers, 1985). This in turn affects their mode of communication in ways that may be expected to flout the cooperative principle.¹ Dawkins and Krebs (1978) proposed that animal signals are often attempts to manipulate the behavior of receivers to the signaler's advantage. This would make signaling the product of an evolutionary arms race between strategic signalers and skeptical receivers, who in turn attempt to deduce the state of signalers to their own advantage. At an equilibrium under natural selection, however, it is likely that both signaler and receiver benefit from the exchange; otherwise the signaling system should cease to exist (Maynard Smith & Harper, 2003). Therefore we should expect a longstanding system of communication among living things to reflect a complex mixture of cooperation *and* conflict.

¹ See Kasher (1977), Sampson (1982), and Sperber and Wilson (1986) for other analyses of indirect speech that do not depend on the cooperative principle.

The second insight is that relationships among conspecifics are expected to fall into discrete kinds, in which the dyads follow qualitatively different sets of strategies depending on their genetic relationship, history, sex, and relative strength. These biologically significant relationship types include dominance, mutualism, parenting, reciprocal altruism, short-term mating, and long-term pair bonding. The hypothesis that humans differentiate their relationships into different kinds has been called *relational models theory* (Fiske, 1991, 1992) and *relationship-specific social psychology* (Wilson & Daly, 1997). It has profound implications for communication, suggesting that the tensions introduced into a relationship by a speech act may not fit along a single continuum (e.g., of face threat or equity imbalance) but may involve uncertainty over which of several possible models of relationship should be in force, each prescribing a different degree and kind of cooperation.

Motivated by these considerations, we have sketched a theory that explains off-record indirect speech in terms of the strategies of speakers in negotiating relationship types under conditions of uncertainty (Pinker, Nowak, & Lee, 2008), which game theorists call *identification problems* (Schelling, 1960). Rather than presupposing full cooperation between speaker and hearer, the theory posits that indirect speech is deployed to negotiate whether and what kind of cooperation should be in effect between them. The logic of this *strategic speaker theory*—that indirect speech supplies plausible deniability in identification problems—is most easily explained with scenarios in which the speaker's expected costs and benefits can be quantified. Such a scenario occurs, for example, when a motorist detained by a police officer contemplates a bribe to evade a traffic ticket. A simple game-theoretic model, which formalizes the intuition of "plausible deniability," specifies the conditions under which indirect speech is the optimal solution for a speaker in such a situation.

Crucially, the same game-theoretic logic applies to purely social situations, such as making a sexual overture or bribing a maitre d', where there may be no legal risks or quantifiable incentives. We suggest that the negotiation of which relational model should be in force defines a payoff matrix that is isomorphic to the one governing the bribe scenario, but where the costs are emotional rather than financial or legal. Indirect speech allows plausible deniability of a breach of a relationship type and thus avoids the aversive social and emotional consequences that would be triggered by such a breach.

Finally, the strategic speaker theory, when extended to higher orders of deniability, can explain why indirect speech is used even when there is little uncertainty in the speaker's intent or the hearer's values regarding the relationship switch. By connecting the qualitative distinctions among relationship types with the digital nature of language, the theory explains why a thinly disguised request that fools no one is still more socially acceptable than the same request expressed baldly. In this article, we develop the theory in full, work out its psychological implications, and report several studies that test its predictions.

Part 1: Indirectness as a Distinct Strategy From Politeness

We motivate the need for the theory of the strategic speaker by showing that pure-cooperation theories, though successful in explaining politeness and on-record indirect speech, are less success-

ful in explaining off-record indirect speech. This extends the conclusions of earlier studies finding that off-record indirect speech does not seem to be an extreme form of politeness but serves largely different purposes (Dillard et al., 1997; Holtgraves & Yang, 1990).

In politeness theory the magnitude of any potential face threat is a monotonically increasing function of three factors: (a) the *social distance* between speaker and hearer, (b) the *power* of the hearer relative to the speaker, and (c) the *degree of imposition* inherent in the speaker's request. Increases in any of these factors lead speakers to use politeness strategies to nullify the resulting face threat. As mentioned, politeness theory posits four strategies that fall along a continuum: positive politeness (sympathy, concern, or camaraderie); negative politeness (respect or deference); and indirectness (the use of a wording that does not literally amount to a request), further subdivided into on-record (pro forma or clichéd) and off-record (novel and oblique) requests. A minor face threat might warrant only positive politeness, whereas a serious threat might need to be cloaked in indirect speech. According to this logic, indirect speech acts are more polite when they are off record because only then can the speaker plausibly deny that he has made a request.

We now present a study which shows that people sometimes favor direct (but polite) requests, even when social distance, power gap, and degree of imposition are manipulated to levels comparable to those in scenarios favoring indirectness. These results indicate that politeness and indirectness do not reside on the same scale but are rather distinct mechanisms elicited by different types of social encounters.

Experiment 1

Politeness theory posits that once a summed face threat crosses a certain threshold, speakers making requests should shift their preference from being direct (but polite) to being indirect. We test this prediction by manipulating the three face-threat factors and showing that they indeed affect politeness but do not push speakers into indirectness. Conversely, social scenarios that pose identification problems (in particular, about whether a hearer is willing to switch the qualitative nature of the relationship) elicit indirectness rather than politeness.

Method

Participants. Participants filled out a questionnaire hosted by Amazon Mechanical Turk (www.mturk.com) at home without any supervision by an experimenter.² Participants had to be at least 18 years old and fluent English speakers. One hundred twenty-eight participants began the questionnaire; 114 completed it. Participants were given \$5 as compensation.

Materials. The questionnaire consisted of four fictional scenarios: (a) a man asking a woman for sex at the end of a date (seduction), (b) a driver attempting to bribe a police officer in order to avoid a ticket (bribe), (c) a professor threatening a talented student with the loss of a scholarship if she does not work in his lab

² Materials for all experiments are available from the authors upon request.

(threat), and (d) a new employee at a finance firm asking a coworker or supervisor for help with a difficult statistical analysis (favor). All four scenarios are given in Appendix A.

The favor scenario was designed to allow for straightforward manipulations of the face-threat variables. Each participant saw eight distinct versions of this scenario in a within-participants design. The eight cells resulted from crossing three factors with two levels each: (a) whether the hearer is another new employee or the speaker's supervisor (a manipulation of power), (b) whether the hearer was the speaker's roommate in college for 3 years or someone whom the speaker hardly knows (a manipulation of social distance), and (c) whether the speaker needs 10–15 min of the hearer's time or more than 3 hr (a manipulation of imposition).

Each scenario was followed by several questions. The first three questions asked participants to rate, on a scale from 1 to 5, the following aspects of the scenario: (a) the extent to which one character was in a position of authority over the other, (b) how socially distant the characters were, and (c) how much of an imposition it would be to grant the speaker's request. The latter question gave participants a *don't know* option in addition to the ratings from 1 to 5.

Participants were then given five different speech acts and asked to rate, on a scale from 1 to 7, how likely it would be for the speaker to use each one to convey the request. The five speech acts were designed to be (a) blunt, (b) positively polite, (c) negatively polite, (d) somewhat indirect, and (e) very indirect. Within each scenario participants always rated the speech acts in this order. The positively and negatively polite speech acts were worded in accordance with the diagnostic features documented by P. Brown and Levinson (1987). The blunt speech act employed a bare minimum of politeness. The blunt and polite speech acts were all unambiguous regarding the nature of the speaker's request. In contrast, the two off-record indirect speech acts did not literally amount to requests. Here we give the speech acts used in the favor scenario:

- “Please help me out with this analysis.” [blunt]
- “So, [hearer], one workaholic to another. I, uh, was thinking that it would be really good for the whole company if I got this report done on time. Could you please help me with it?” [positively polite]
- “I'm really sorry to bother you, and I wouldn't ask this if it wasn't hugely important. But do you think it might be possible to step through this analysis with me? It would be a real life-saver.” [negatively polite]
- “I really admire you, [hearer]. You have the perfect background for this. I wish someone could have told me in school that this stuff would be really invaluable in my work.” [somewhat indirect]
- “Oh, I can't believe this. I'm probably going to miss the deadline because of this problem.” [very indirect]

After rating each of the speech acts, participants had to select the one that they thought was closest to what the speaker would actually say, with the option of rephrasing the closest selection in their own words if desired. This was to ensure that the selection would not be overly affected by the specific wordings that we chose.

Procedure. The 11 scenarios (seduction, bribe, threat, and eight cells of favor) were arranged in 32 distinct orders. The different versions of the favor scenario were always adjacent. The favor scenarios as a group were rotated along with the bribe, seduction, and threat scenarios in a Latin square. For each of these four possible orderings, the favor scenarios were rotated among themselves in a Latin square. There were eight such orderings of the favor scenarios, leading to 32 orders in total. A maximum of four participants were allowed to complete each order.

Results

We first tested the validity of participants' ratings of face threat by seeing whether each rating scale in the favor scenario was most strongly influenced by the corresponding manipulation in the design of the materials. Ratings of the power gap were influenced most strongly by our manipulation of the power gap (*effect size* = 1.6 Likert points, 4 times that of the next strongest manipulation), though they were somewhat influenced by the other manipulations as well ($p < .01$). The ratings of social closeness were significantly affected only by the manipulation of social closeness itself ($p < .001$). The ratings of imposition were most strongly influenced by the manipulation of imposition (*effect size* = 1.1 Likert points, 3 times that of the next largest manipulation), though they were also affected by the other manipulations ($p < .001$). Thus the ratings of the face-threat variables are sensitive to the underlying constructs.

We next examined whether manipulations of the face-threat factors in the favor scenarios had the effects on language predicted by politeness theory. We estimated the effect of each face-threat manipulation on participants' endorsement of each speech act by restricted maximum likelihood (REML), treating participant variability as a random effect (see Table 1). Each entry in the last three columns reflects the effect of increasing the magnitude of the face-threat factor on the attractiveness of the speech act in that row.

The blunt speech act was surprisingly popular. In fact, when all face-threat factors theoretically favored minimal politeness, the blunt and negatively polite speech acts received nearly equal mean ratings. As the face-threat factors were manipulated to favor more politeness, the ratings of these two speech acts diverged in precisely the manner predicted by politeness theory, with the negatively polite speech act becoming more popular and the blunt speech act less so. The effects of manipulating degree of imposi-

Table 1
Results of Experiment 1: Within-Participants Manipulation

Speech act	Intercept ^a	Power gap	Social distance	Imposition
Blunt	4.52	-0.37***	-1.17***	-0.03
Positively polite	3.11	-0.25**	-0.09	0.03
Negatively polite	4.58	0.20*	0.54***	0.12
Somewhat indirect	3.51	0.28**	-0.02	0.26**
Very indirect	3.67	-0.46***	-0.77***	0.26**

^a The average Likert rating of how likely it would be for the speaker to use the speech act when power gap, social distance, and degree of imposition were all set at the lowest level.

* $p < .05$. ** $p < .01$. *** $p < .001$.

tion were not significant for either the blunt or negatively polite speech act, but the signs of these effects were correctly predicted.

The positively polite speech act was surprisingly unpopular and relatively insensitive to manipulations. It may be that the presumption of friendliness and shared goals inherent in positive politeness strikes participants as manipulative or presumptuous in this kind of situation. This confirms earlier suggestions that the positive and negative politeness strategies may not always be ordered along a single dimension (Holtgraves, 2002; Lim & Bowers, 1991; K. Tracy, 1990).³

As predicted by politeness theory, the somewhat indirect speech act received higher ratings as power gap and degree of imposition were increased. However, the ratings of this speech act never approached those given to the negatively polite speech act; participants consistently indicated that the negatively polite speech act was most appropriate in this scenario. The very indirect speech act became disfavored as the face-threat factors were manipulated to favor politeness. The participants' forced-choice responses were consistent with their ratings: The negatively polite speech act was the most popular choice in all eight cells, its share ranging from 34% in the cell calling for the least politeness to 60% in the cell calling for the most. This dispreference for indirect speech suggests that politeness theory, while accounting for the use of negative politeness in conversation, fails to account for the use of off-record indirect speech.

The second critical test comes from the seduction, bribe, and threat scenarios, which embody identification problems. Here we found a very different pattern from the preference for negative politeness holding in the favor scenario: Participants favored the two indirect speech acts over the three direct ones. To quantify this effect, we scored each forced-choice response as 1 if it endorsed one of the indirect speech acts and 0 otherwise. When these responses in each of the seduction, bribe, and threat scenarios were compared to the responses in the cell of the favor scenario where all face-threat factors favored maximal politeness, McNemar's test for paired dichotomous data showed that participants were more likely to endorse an indirect speech act in all three nonfavor scenarios—seduction: $\chi^2(1, N = 114) = 57.1, p < .001$; bribe: $\chi^2(1, N = 113) = 15.3, p < .001$; and threat: $\chi^2(1, N = 114) = 52.7, p < .001$. In the seduction scenario, 91% of participants opted for indirectness; in the bribe scenario, 58%; in the threat scenario, 86%. In the different cells of the favor scenario, in contrast, indirectness garnered between 21% and 34% of the responses. Responses to the Likert rating scales showed the same pattern.

Of course, it is crucial to show that the preference for indirect speech in the seduction, bribe, and threat scenarios was not just a consequence of their posing greater degrees of face threat than the favor scenario. It was not. Ratings of the three face-threat factors for these scenarios were comparable to or lower than those for the cell of the favor scenario favoring maximal politeness (see Table 2). As a rule of thumb, any pairwise difference within a column of Table 2 exceeding 0.3 is statistically significant at $\alpha = .05$. The only scenario that equaled or exceeded the favor scenario in ratings of the face-threat factors is the bribe scenario. However, in all three of the nonfavor scenarios, participants overwhelmingly favored indirect over direct requests, and out of the nonfavor scenarios it was in fact the bribe scenario that was most likely to elicit endorsement of a direct request.

Table 2
Results of Experiment 1: Degree of Face Threat Across Scenarios

Scenario	Power gap	Social distance	Imposition ^a
Favor ^b	4.29	4.22	4.14
Seduction	1.73	2.64	4.49
Bribe	4.68	4.49	4.08
Threat	3.89	2.92	3.60

^a Participants who selected *don't know* were ignored in calculating these means. ^b The cell in which power gap, social distance, and degree of imposition were all set at the highest level.

One other finding highlights the mismatch between politeness theory and scenarios involving social identification problems. Intuitively, the question about the degree to which a sexual proposition from a man to a woman requests an "imposition" is ill-formed. If the woman doesn't want to have sex with the man, the prospect is so aversive that "imposition" seems an inadequate term; if she does want to, then it is not an "imposition" at all. To a lesser extent this applies to other scenarios involving identification problems: A police officer may be corrupt (and accept a bribe) or honest (and rebuff it); the target of a threat may find it expedient to accede to the threat, or he may defy the threatener (either to put the onus on him to enforce it or to deter similar threats in the future). For this reason, when asking the participants to rate the degree of imposition, we gave them the option *don't know*. In the different cells of the favor scenario, the proportion of participants choosing the *don't know* option ranged from 2% to 8%. In contrast, in the seduction scenario, 48% of the participants chose the *don't know* option; in the bribe scenario, 12% chose it; and in the threat scenario, 19% chose it. The difference between the number of *don't know* responses in the maximal cell of the favor scenario and the numbers in the seduction and threat scenarios were significant by McNemar's test, $\chi^2(1, N = 114) = 38.9, p < .001$, and $\chi^2(1, N = 114) = 5.76, p < .05$, respectively.

Discussion

We confirmed the predictions of politeness theory on the relative preference between blunt and negatively polite (deferential) speech in asking favors. Increasing the three face-threat factors (power gap, social distance, degree of imposition) led participants to shift their predictions of the wording of a speaker's request for a favor from bluntness to negative politeness. However, no matter how high the ratings of the face-threat factors became, participants did not make the further shift from the negatively polite speech act to either of the indirect speech acts. This was despite the fact that ratings of the face-threat factors for some of the cells matched or exceeded the corresponding ratings in the seduction, bribe, and

³ In a well-known routine (see Raggedclown, 2006), the Irish comedian Dave Allen notes that terms of solidarity like *chum*, *amigo*, and *buddy* are often used aggressively, as in *Listen to me, pal!* and *All right, mate, you want it, you can come and get it!* Presumably this is because a presumption of intimacy when none exists is seen as a perquisite of a dominant individual.

threat scenarios, in which participants strongly favored the indirect speech acts.

We conclude that these scenarios differ along dimensions other than the three face-threat factors. We suggest that the appeal of the *don't know* option when rating degree of imposition reveals the critical difference. When a less competent person asks a more competent person for a favor, he or she is always making some imposition, however small. But in seductions (and other interactions where the values of the hearer are not known to the speaker), the request does not have this character. For example, at the end of a pleasant evening, the woman might be perfectly willing to have sex with the man, if only she can be certain of what exactly he is requesting.

In the favor scenario, then, the values of the hearer are "identified" in that the speaker knows the request poses some imposition. In the other scenarios, however, different hearers might hold qualitatively different attitudes toward the speaker's request. The results of this study indicate that the use of indirect speech in conditions of uncertainty about the degree and nature of cooperation desired by the speaker requires an explanation beyond that provided by politeness theory and other pure-cooperation theories.

Part 2: Plausible Deniability as a Solution to the Identification Problem

To illustrate the appeal of indirect speech in identification problems, we begin with a subtype of conversation in which the costs and benefits may be quantified. Consider a motorist who has been pulled over for speeding and is considering two options: accepting the ticket or offering a bribe to the police officer. Which offers the better expected payoff? The answer is that it depends on the *values* of the officer he is facing. Bribery is a criminal offense which is punished much more harshly than a traffic violation, so if the officer is a scrupulous enforcer of the law, a bribe would lead to a more severe penalty for the driver than the ticket. He thus faces an identification problem: There are two types of hearers, cooperators and antagonists. An attempted initiation of a joint endeavor with an antagonist will backfire to result in the worst possible outcome, and the speaker cannot tell the two types apart.

Suppose that the driver in the bribe scenario were constrained to obey Grice's (1975) cooperative maxims to be informative, succinct, truthful, relevant, and clear. Such a driver would therefore offer the bribe in direct speech, as an overt quid pro quo. The form of the identification problem faced by such a driver is depicted in the first two rows of the payoff matrix in Table 3. The driver can either accept the ticket or offer a direct bribe such as *If you let me go without a ticket, I'll give you fifty dollars*. Given the large potential cost inherent in offering a bribe to an honest officer,

however, the driver might instead refrain from bribing and swallow the smaller cost of the ticket.

But now suppose that the driver is allowed to deviate from the maxims by veiling the bribe in an obscure, ambiguous, or irrelevant comment, such as *So maybe the best thing would be to take care of that here* (represented as the third row of Table 3).⁴ Suppose that a dishonest officer can detect the bribe in the innuendo using mechanisms of implicature documented by linguists (P. Brown & Levinson, 1987; Clark, 1996; Grice, 1975; Holtgraves, 2002; Horn, 2003; Searle, 1975). And suppose that an honest officer, hearing the same innuendo, could not make a bribery charge stick in court by the high standard of proof beyond a reasonable doubt. Then the indirect speech may allow the driver to combine the large benefit of bribing a dishonest officer with the relatively small cost of submitting to an honest one.

Though the appeal of indirect speech in attaining plausible deniability may seem like common sense, more rigorous examination shows that indirectness is not advantageous across the board but only with a particular configuration of expected payoffs. The circumstances under which indirect speech is optimal can be derived from a simple game-theoretic model (Pinker et al., 2008). Let y denote the expectation of the cost borne by the driver. This cost depends on: (a) q , the proportion of officers who are honest; (b) c_0 , the cost of the bribe; (c) c_1 , the cost of the ticket (which must be greater than the cost of the bribe, or else it would never pay to bribe); (d) c_2 , the cost of an arrest for bribery (which must be greater than the cost of the ticket, or else it would always pay to bribe); and (e) p , the probability that an officer will interpret the driver's statement as an attempted bribe and act accordingly—accepting the bribe or arresting the driver, depending on whether the officer is cooperative (corrupt) or antagonistic (honest). p increases monotonically with the directness of the statement, d , which is the crucial linguistic variable. It reflects the number of inferential steps that must be carried out in the implicature to get from the literal content of the sentence to the meaning intended by the speaker. Critically, the linguistic variable of directness must correspond to the decision-theoretic variable of the probability that a signal reflects one state of affairs or another.

It can be shown that if corrupt and honest officers use the same decision function when acting on a driver's statement, indirect speech is never optimal. There exists a critical proportion of officers who are honest, $q_c = (c_1 - c_0)/(c_2 - c_0)$, such that if $q > q_c$, the optimal strategy for the driver is not to attempt a bribe at all ($p = 0$), and if $q < q_c$, the optimal strategy for the driver is to offer a direct bribe in clear and unmistakable terms ($p = 1$). Under these assumptions, therefore, the model predicts that indirect speech should never be attractive to a speaker. If the proportion of antagonistic listeners is high, he should remain silent; if it is low, he should offer the bribe overtly.

There is, however, a set of conditions in which indirect speech is optimal: when the two types of hearers *differ* in their probabilities of responding to the speaker's first move. Suppose that a corrupt officer must meet a threshold of certainty (say 80%) that a driver's utterance is an attempted bribe before he will remove the cash from a proffered wallet and send the driver on his way. Now,

Table 3
Payoff Matrix for Speeding Motorist

Driver's strategy	Type of officer	
	Dishonest officer	Honest officer
Don't bribe	Traffic ticket	Traffic ticket
Direct bribe	Go free	Arrest for bribery
Indirect bribe	Go free	Traffic ticket

⁴ The example is taken from a scene in the movie *Fargo* (Coen & Coen, 1996).

if an honest officer must meet an even higher threshold (say 99%) before he will arrest the driver for bribery, a rational driver should pitch his utterance at a directness level intended to induce a level of certainty in the hearer somewhere between 80% and 99%. (Of course, the driver need not be aware of the rationale behind his choice of words.)⁵

The critical assumption is that over some range of directness, a hearer antagonistic to the speaker sets a higher threshold for action than one who is completely cooperative. This is plausible in the case of an attempted bribe, because even if an honest officer believes just as strongly as his corrupt counterpart that a given utterance is a bribe, he might be less willing to act on his belief because of the doubtful prospects of a successful prosecution, the risk of a lawsuit for wrongful arrest, the additional paperwork, and other costs. These contingencies could easily be modeled in the payoff matrix for the hearer.

Note how this model differs from the traditional assumption that indirect speech is governed by pure cooperation: The strategic speaker is using indirectness not to work with the hearer in pursuing a joint goal (in this case, to enforce the law) but rather to subvert that goal.

The theory's central concept of deniability is related, but not identical, to the concept of *defeasibility* in the linguistic literature on implicature (Grice, 1975; Horn, 2003). An implicature is said to be defeasible if it is not logically entailed by the utterance and can be canceled or suspended without contradiction. For example, *Some men are chauvinists* naturally leads listeners to infer that the speaker also means that some men are not chauvinists. But that implicature is not logically entailed and can be canceled, as in the sequence *Some men are chauvinists; indeed, all are*. (Similarly: *They got married and had a baby, but not in that order*.) Though deniability requires defeasibility, the reverse is not true; defeasibility does not require that the speaker have a self-interested reason to cancel the implicature with antagonistic hearers but not with cooperative hearers. In the standard examples in linguistics, defeasibility is exploited purely for rhetorical purposes, and the utterance is meant to be understood in the same way by all hearers.

Plausible deniability is also distinct from the maxim "provide options" in theories of politeness (P. Brown & Levinson, 1987; Lakoff, 1973), which underlies a range of constructions of deferential politeness such as *If you please* and *Do you think you might*. This maxim is intended to ease the face threat of a request directed to a cooperative hearer, with no strategic consequences; it is not a strategy designed to protect the interests of the speaker when addressing a potentially antagonistic hearer. Indeed, Experiment 1 gave participants the choice between a provide-options construction (*Do you think it might be possible*) and an indirect request and found that the two are preferred in complementary circumstances.

The model of a strategic speaker has testable implications. First, the indirectness of speech should be perceived not just as a social gesture that redresses the hearer's face (such as the use of constructions like *Please*, *Can you*, and other politeness reflexes) but as one that monotonically affects his interpretation and decision. That is, people should sense that the abstract linguistic variable of directness, which underlies an unlimited range of wordings, corresponds to the hearer's probability of action, and they should deploy that correspondence strategically, so as to minimize their quantifiable costs according to the payoff matrix implicit in the situation. The use of directness should thus be sensitive to vari-

ables that go into that payoff matrix, such as the proportion of cooperative and antagonistic speakers in the population and the costs of consummating or failing to consummate a cooperative relationship, but that are omitted from theories of indirect speech that assume the relationship between speaker and hearer to be universally cooperative. Thus directness should vary even when the face-threat variables emphasized in politeness theory (power, distance, and imposition) are held constant. If, in contrast, speakers calibrate their linguistic directness only to face threat, not to payoffs and probabilities, it would undermine the hypothesis that indirectness is deployed strategically in cases where cooperation is uncertain.

The second critical prediction is that when speakers do favor indirect speech, they should also perceive that cooperative hearers (in this case, dishonest cops) have lower thresholds of action along the directness continuum than uncooperative hearers (in this case, honest cops). If there is no such difference, then strategic speaker theory would clearly be falsified, because it is only under those circumstances that indirect speech is optimal.

Experiment 2

To test the strategic speaker theory in a scenario in which costs, benefits, and decisions can be specified precisely, we had participants imagine that they were tempted to bribe a police officer. The theory does not, of course, claim that speakers develop the use of indirect speech specifically so that they can bribe officials with cash, but rather that they easily generalize the strategies of indirectness they use in everyday conversation to these more specialized (and, for our purposes, quantifiable) situations.

Method

Participants. Fifty-one fluent English speakers were recruited from the Harvard undergraduate student body to fill out an online questionnaire without compensation; 26 completed it.

Materials. The questionnaire was hosted on www.surveymonkey.com and was taken at home without supervision. Participants were asked to imagine themselves on a road trip through a fictitious former republic of the Soviet Union. The first part of the questionnaire was intended to familiarize participants with the scenario and payoff structures.

An excerpt from a guidebook informs participants that because corruption and bribery are rampant in this part of the world, a traveler must understand how to act in an encounter with a police officer. The guidebook then gives four versions of a scenario in which the participant imagines being pulled over by an officer who claims, falsely, that the participant was speeding. Each version differed in the values of the variables predicted to be relevant to

⁵ Pinker et al. (2008) noted that the same result holds if the hearers obey a continuous function for the probability of their decision, rather than a step function with a discrete threshold. The model can also be extended to a discourse of escalating propositions along the directness continuum which probe the hearer's decision function in smaller increments (e.g., "What a beautiful morning. I'm very sorry for speeding. I know I'll have to pay for my mistake. I admire officers doing their duty. Can I make a contribution to the policeman's benevolent association? Is there some way we could avoid the paperwork and settle it here?").

the optimal choice of speech act: (a) the proportion of police officers in the area who are honest, (b) the cost of the expected bribe, (c) the cost of the ticket, and (d) the cost of being arrested for bribery (see Table 4). In each case the scenario concludes by asking participants to consider how they would phrase an offer of a bribe to extricate themselves from the situation.⁶

Participants selected one speech act from a menu of five options in response to each familiarization scenario and received one of three possible kinds of feedback on the outcome. The choice of feedback depended on the predicted outcome of the participant's response given the parameter values of the scenario:

- You encountered a willing officer, but he still wrote you a ticket.
- You encountered a willing officer. He takes your bribe and drives off.
- You encountered an unwilling officer. He punishes you for attempting to bribe an officer, and you have to pay [the amount of the fine].

Participants were told to imagine themselves with \$2,000 at the start of these encounters and to keep track of their total after each successive loss (ticket, successful bribe, punishment for attempted bribery). Fields in the questionnaire were provided for participants to type in their totals.

At this point the experimental trials began. Participants were told that they would visit eight different counties on a road trip, each leading to an encounter with a policeman whose status as a corrupt officer (cooperator) or honest officer (antagonist) was unknown.

Eight combinations of values were used in the experimental trials, summarized in Table 4. In half of the conditions, the ratios of bribe:ticket:fine were 1:9:10; in the other half, they were 1:2:10. All else being equal, the ticket inflicts a greater loss on the driver in the former conditions, so these should tilt participants toward more direct speech acts. The scalar differences do not affect these predictions. In addition, a higher proportion of officers who are honest should tilt participants toward less direct speech acts. Participants were encouraged to pay attention to the numbers, which were highlighted in bold.

After each experimental trial, participants were shown the following four speech acts in randomized order:

Table 4
Levels of Experimental Factors in Experiment 2

Trial ^a	% honest (<i>q</i>)	Bribe (<i>c</i> ₀)	Ticket (<i>c</i> ₁)	Fine (<i>c</i> ₂)
Familiarization 1	.75	\$40	\$280	\$400
Familiarization 2	.25	\$40	\$280	\$400
Familiarization 3	.25	\$10	\$280	\$290
Familiarization 4	.50	\$50	\$350	\$500
Experimental A	1	\$50	\$450	\$500
Experimental B	1	\$20	\$40	\$200
Experimental C	0	\$40	\$80	\$400
Experimental D	0	\$70	\$630	\$700
Experimental E	.90	\$50	\$100	\$500
Experimental F	.10	\$30	\$60	\$300
Experimental G	.90	\$30	\$60	\$300
Experimental H	.10	\$40	\$360	\$400

^a The experimental trials were not administered in the listed order but according to a Latin square.

• "I'm very sorry, officer. If I give you a fifty, will you just let me go?" [overt]

• "I'm very sorry, officer. But I'm actually in the middle of something right now, sort of an emergency. So maybe the best thing would be to take care of this here . . . without going to court or doing any paperwork." [nearly overt]

• "I'm very sorry, officer. I know that I'll have to pay for my mistake." [indirect]

• "I'm very sorry, officer. I've really learned my lesson." [very vague]

Participants were asked to allot 100 "probability points" among these four speech acts, each allotment corresponding to the strength of the participant's liking for the speech act. The wording of each of these speech acts was distinct from all of those used in the familiarization trials; this was to ensure that participants were not merely learning to associate rewards and punishments with particular linguistic formulas or constructions. Participants were instructed not to carry out the allotment task unless they were certain that they understood it. No feedback was provided after participants indicated their judgments.

The speech acts used in the experimental trials were chosen in the following way. In a pilot study, 30 Harvard University students were asked to rate the directness of eight wordings of a bribe on a scale from 0 to 100. From these eight wordings, we selected four that showed separation from each other in a parallel boxplot ($p < .01$ by the Wilcoxon signed rank test) and designated them as *very vague*, *indirect*, *nearly overt*, and *overt* accordingly. (The two wordings of intermediate directness were the same as those used in the Experiment 1 bribe scenario.)

As an additional check that participants interpreted a speaker's directness as a cue for probability of intent, we asked them to estimate, for each of the four wordings, the probability that a corrupt officer would interpret it as an attempted bribe and carry out the exchange. And to test the critical prediction about differing thresholds for cooperative and antagonistic hearers, the questionnaire asked for the probability that an honest officer would respond to that sentence by arresting the driver for attempted bribery.

Procedure. All participants completed the familiarization trials in the same order. Before the start of the experimental trials, participants were asked to select the top choice in a drop-down menu displaying the numbers 1 through 8 in a random order. The participant was then directed by a link to a sequence of experimental trials corresponding to that number. These sequences were generated by a Latin square.

Results

We summarized each participant's responses to the experimental trials by treating the allotment of points as a probability distribution over the four sentences. We then assigned the sentences the values 1, 2, 3, and 4 in order of increasing directness and calculated the expectation. This quantity can be interpreted as the expected directness of the participant's bribe. We treated this as

⁶ Pilot testing revealed a tendency for participants to favor the most indirect speech act possible across the board. To reduce this floor effect, we added context asking participants to imagine that on the previous night they encountered the officer off-duty in a bar.

the outcome variable in mixed linear models estimated by REML, incorporating participant variability as a random effect and the relative cost of the ticket and the proportion of officers who are honest as fixed effects. The relative cost of the ticket was dichotomized as 1/0.

Three participants gave the same stereotyped response (equal weight to all options or all weight on one) to four or more consecutive experimental trials and were removed from the dataset, leaving 23 for analysis. At least two participants completed each of the cells.

Figure 1 displays the cell means. The fixed effect of ticket cost was statistically significant, $t(18) = 2.19, p < .05$. The fixed effects of the proportion of officers who are honest, relative to a baseline of no honest officers, were also significant, $t(18) = 4.77, p < .001$; $t(18) = 14.25, p < .001$; $t(18) = 15.88, p < .001$. These findings are consistent with the predictions of the strategic speaker model.

Figure 2 shows the perceived probabilities that the two types of officers will respond to speech acts varying in directness, a test of the hypothesis that indirect speech is used strategically. Setting aside a couple of outliers, we see that the direct offer was believed to lead deterministically either to an arrest or to a consummated transaction. For the three speech acts of varying indirectness, participants tended to believe that honest officers are less likely to arrest the driver for attempted bribery than corrupt officers are to accept the implicit offer. These tendencies were statistically significant by the signed rank test ($p < .01$). The results confirm that the linguistic variable that differentiates “vague” from “overt” sentences (as judged by the participants in the pilot study) is related to the perceived probability that a hypothetical speaker will interpret and act on the sentence as an actual request, with predicted asymmetries between cooperative and antagonistic speakers.

Discussion

This study confirms two predictions of the strategic speaker model. The first is that the choice of directness in speech is affected not just by the traditional variables of power, status, and imposition but by the payoff structure inherent in the different

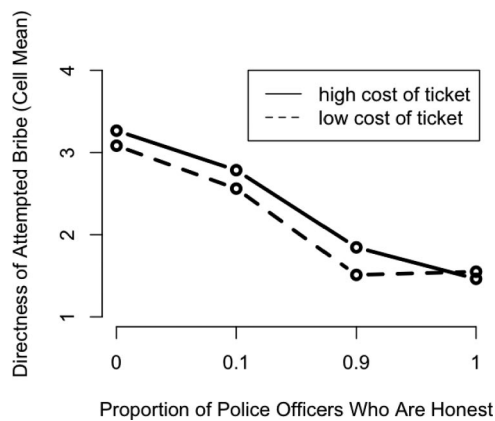


Figure 1. Mean preferred directness of an attempted bribe in Experiment 2. Increments along the x-axis are merely ordinal.

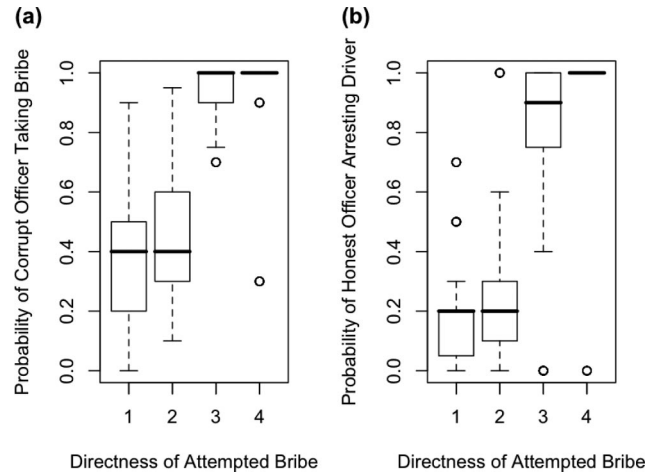


Figure 2. Parallel boxplots of Experiment 2 participants’ estimations of probabilities that (a) corrupt officers will interpret the speech acts as bribes and take the money and (b) honest officers will interpret the speech acts as attempted bribes and arrest the driver.

possible interpretations and actions of the hearer. Note that participants were never told to attend to the directness of the response options; nor did the familiarization trials expose them to those wordings. They had to infer that the directness dimension was relevant to the payoffs and extend it to their choice of new wordings. The finding that participants applied linguistic directness to maximizing payoffs under uncertainty is consistent with the hypothesis that the directness of a speech act is an adaptive response to a game-theoretic identification problem.

The second confirmed prediction of the strategic speaker model is that people assume that the threshold for appropriate action set by an antagonistic hearer to be higher than the one set by a cooperative hearer. This condition allows an off-record indirect speech act to be the optimal solution to an identification problem.

Of course, scenarios in which the costs and benefits of a hearer’s interpretation can be quantified are not the most common context in which indirect speech is used. In everyday life, we would expect that intangible social and emotional costs, not just money, define the payoff matrices that call for indirect speech. One finding in this experiment underscores that expectation. Figure 1 shows that when conditions favored a bald proposition (0% policemen honest), participants often did not conform to the seemingly rational response of placing their entire probability mass on the most direct speech act, but rather opted to be relatively vague. This bias is not attributable to experimental noise or a tendency of the participants to avoid extreme responses. When the payoffs militated toward the opposite extreme and ruled out an attempted bribe entirely (100% policemen honest), responses were bunched more tightly at the vague end of the scale. Nothing in the model (at least when implemented with dollar values) predicts this asymmetry. Participants must have felt that there is some emotional cost incurred in a botched encounter with a police officer over and above any state-enforced penalties. We now turn to the intangible social and emotional costs that are the more typical payoffs for the strategic speaker.

Part 3: Relationship Negotiation

Though everyday social life may lack the formal regulations that allow plausible deniability to be quantified, it often presents similar contingencies in the currency of emotional discomfort. Consider a case that is close to bribing a police officer: bribing a restaurant maitre d' in order to be seated immediately despite having no reservation. No one has ever been fined or imprisoned for attempting to bribe a maitre d', yet in a telling real-life example, a writer given this assignment by a magazine editor reported that he always tendered the bribe indirectly (*Is there any way to shorten my wait?*) rather than directly (*I will give you \$20 if you seat me immediately*; Feiler, 2000). Similarly, the common expectation is that a man propositioning a woman after dinner would use an indirect come-on (*Would you like to come up for a cup of coffee?*) rather than a direct one (*Would you like to come up and have sex?*). Once again, no one would lead the man away in handcuffs if the woman were to rebuff his advance, yet he acts *as if* someone might do so, just as in the scenario with the driver and the police officer.

To apply the model of the strategic speaker to cases where tangible costs are absent, we must identify the intangible costs that define the requisite payoff matrix. One possibility is the implicit currency of social equity invoked by Clark (1996), who noted that actions leading to an equity imbalance between speaker and hearer prompt them to redress it with compensating actions or verbal gestures. The problem for this account is that a bribe or a sexual favor offers the opportunity for perfect equity if it is mutually agreed upon, yet merely broaching such a possibility can lead to enormous emotional tension. This suggests that emotional tension is not just a matter of equity imbalances. Not all relationships are governed by equity, and those that are may differ in which resources may be legitimately entered into the balance sheet. We suggest that it is the very process by which two people choose to enter an equity relationship (or not) that gives rise to the relevant emotional costs.

Qualitative differences among relationship types, and the emotional costs of uncertainty regarding which one applies, have been explored in cross-cultural detail by Fiske and his collaborators (Fiske, 1991, 1992; Fiske & Tetlock, 1997; Haslam, 1994a, 1994b). They proposed that humans use just four discrete *relational models* to govern their interactions. Each is defined by a distinct calculus for the apportionment of resources, and each has a distinct evolutionary basis. Humans assign every relationship and resource to one of the four models, but the assignments can vary with the culture and sometimes can be renegotiated dynamically by individual dyads.

The first model is *communal sharing*. Arising from the evolutionary forces of *kin selection* (Hamilton, 1964; Maynard Smith, 1964) and *mutualism* (Tooby & Cosmides, 1996), it applies to relationships among family members, spouses, and close friends with a strong overlap of interests. Partners in a relationship of communal sharing bond through signals of physical union such as bodily contact and common meals or rituals, and the hallmark of such relationships is the free sharing of resources with minimal regard for equity balances.

The second is *authority ranking*, which governs individuals at different levels of a dominance hierarchy. Signaled by signs of size, strength, and priority, it empowers a dominant individual to

seize resources at will (Maynard Smith, 1982), again with little concern for equity.

The third model is *equality matching*, which implements the evolutionary logic of *reciprocal altruism* (Trivers, 1971) and other ways of apportioning resources on an equitable basis. It governs relationships among acquaintances, neighbors, and trading partners, is defined by more-or-less explicit mechanisms of ensuring equity, and may be negotiated by verbal contracts.

Fiske's taxonomy includes a fourth model called *market pricing*, which applies to transactions among buyers, sellers, lenders, and middlemen in a modern monetary economy. Unlike the other three relationship types, market pricing is not a universal feature of human societies (D. E. Brown, 1991) and does not come naturally even to participants in a complex modern economy (Caplan, 2002). People instead tend to regress toward equality matching, with its face-to-face exchanges of tangible quantities. For this reason we do not distinguish equality matching and market pricing.

When two people perceive that one of these relational models applies to their interaction in a given context, each tacitly accepts the designated kinds of transactions as socially legitimate. Transactions appropriate to other models, in contrast, elicit feelings of awkwardness or, if they are deliberate and sustained, moralistic anger (Fiske & Tetlock, 1997; Tetlock, 2003). For example, a husband can pick an hors d'oeuvre from his wife's plate or vice-versa (communal sharing), but an employee cannot snatch an hors d'oeuvre from the plate of his supervisor (violation of authority ranking). Similarly, the sale of a car or house can feel perfectly comfortable between strangers or acquaintances (equality matching or market pricing), but it becomes awkward if conducted between close friends (a violation of communal sharing). As these examples suggest, within a given culture relational models are not assigned to dyads across the board but may be differentiated according to the resource. A professor and graduate student may implement communal sharing in most social resources (e.g., priority in line in a cafeteria) but authority ranking in professional resources (e.g., use of lab equipment); platonic opposite-sex friends may obey communal sharing for most resources other than sexuality.

When one party breaches the relational model currently in force, the result is one of the "self-conscious emotions" such as awkwardness, embarrassment, or shame (Haidt, 2003; J. L. Tracy, Robins, & Tangney, 2007). These are aversive emotions that pain speakers in the present and motivate them to avoid similar affronts in the future. Blushes, restless eye movements, stammering, confusion, and other involuntary signs of shame may thus serve as *honest signals* of the speaker's remorse and concomitant resolve to refrain in the future from similar offenses (Hauser, 1996; Maynard Smith & Harper, 2003). This interpretation is consistent with the fact that displays of shame and awkwardness appear only in the presence of others and tend to be particularly intense in the presence of the wronged individual.

When the violation of a relational model is not accidental or transient but deliberate and protracted, the actions are not just embarrassing but are stigmatized by taboo and may be formally criminalized. This may happen when a person proposes to allot resources that are normally governed by communal sharing or authority ranking according to the rules of equality matching or market pricing instead. Examples include prostitution; extortion; the solicitation of a quid pro quo sexual favor; the sale of votes,

organs, or adoption rights; and buying one's way out of jury duty or military service (Fiske & Tetlock, 1997; Tetlock, Kristel, Elson, Green, & Lerner, 2000). People care intensely about the prescriptions of relational models because these models are the means by which humans agree on how to distribute the resources (material, emotional, and sexual) needed by all members of their communities for survival and reproduction.

We can now characterize the intangible costs that can drive a strategic speaker to indirect speech. Many propositions assume a relational model governing the resource in question: An onerous command assumes authority ranking; an offer to exchange assumes equality matching; a proposition for consensual sex assumes communal sharing; and so on. A speaker resorts to indirect speech when the relational model assumed by the speech act clashes with the model that currently holds between the speaker and hearer, avoiding the risk of awkwardness or shame in the same way that a briber avoids the risk of an arrest.

Consider the writer assigned to bribe a maitre d'. He reported that the assignment filled him with dread as he kept "imagining the possible retorts of some incensed maitre d': 'What kind of establishment do you think this is?' 'How dare you insult me?'" (Feiler, 2000). The reason for the anxiety, we suggest, is that whereas the relationship conventionally assumed by a maitre d' toward diners is one of authority ranking, according to which he seats diners when and where he pleases, the writer was proposing a reciprocity transaction appropriate to equality matching, according to which the maitre d' would be obligated to seat him in exchange for accepting the bribe.

This situation is now isomorphic to the identification problem in the officer–driver scenario, but now the incentives are not monetary or legal but emotional, reckoned in the awkwardness triggered by a mismatch of relational models. Table 5 presents the new payoff matrix. If an honest maitre d' sets a higher evidentiary standard for acting on an attempted bribe, then the optimal strategy for the diner is to employ indirect speech, such as "Is there any way to shorten my wait?"—which is in fact what the magazine writer spontaneously did (Feiler, 2000).

The reason an honest maitre d' might set a higher evidentiary standard for overtly rebuffing a bribe (compared to the one a corrupt one would set for accepting it) is also parallel to the case of the honest officer: His own payoff matrix differs from that of his corrupt counterpart. In particular, he has to anticipate the lost business from the customer and onlookers that would ensue if a diner stoutly and plausibly defended his innocence. In Part 4, we

explore a more general explanation for the higher standard required by antagonistic hearers.

Note that Fiske's theory of relational models helps to explain the overall tendency to avoid directness (superimposed on the strategic use in response to the monetary payoffs) found in Experiment 2. A police officer wields a very strong dominance relationship over a detainee (far greater than that between a maitre d' and a diner). The emotional costs of impugning this relationship may have been enough to discourage some participants from endorsing the blatant offer of a bribe even in the conditions where all officers were said to be corrupt and the financial costs should have militated toward directness.

Part 4: Higher-Order Deniability

Before turning to a further experiment testing the strategic speaker theory, we address an additional puzzling phenomenon surrounding off-record indirect speech. Regardless of whether a rebuffed proposition leads to a tangible cost (Part 2) or an emotional cost triggered by a relational model mismatch (Part 3), the strategic speaker model requires *uncertainty* on the part of the speaker regarding the hearer's values and on the part of the hearer regarding the speaker's intent. But Experiment 2 showed that speakers tended toward indirect speech even when there was no uncertainty about the hearer's values (i.e., when all officers were known to take bribes). And in everyday conversation, many indirect speech acts are so pro forma or transparent as to leave little doubt about the speaker's intent. If there is no uncertainty about either what the speaker means or whether the hearer is on the speaker's side, then how can an indirect speech act minimize the speaker's expected cost? The remaining puzzle, then, is the use of indirect speech acts even when deniability seems to be neither plausible nor required.

We propose that a special property of relational models combines with a special property of language to allow indirect speech to generate higher order kinds of deniability. Fiske and his collaborators argued that relational models are discrete systems and that for any given resource and context, a pair of people conform to only one of them (Haslam, 1994a, 1994b). And it has long been noted that language is perceived as a digital medium: one that conveys information in a discrete, context-independent, all-or-none manner, rather than a blended or graded manner (Pinker, 1994, 1999, 2007). The discreteness of language is apparent in the distinctness of words and phonemes and their arbitrary relationship to meaning: One cannot express an action partway between *batting* and *patting* with a sound partway between *bat* and *pat*, nor can the action of talking while walking be predictably expressed by the sound *twalking*.

More generally, we suggest that people implicitly perceive direct speech as being capable of expressing intentions unambiguously, losslessly, and recursively—each a feature of digital transmission. These features in turn allow indirect speech to provide forms of *higher-order plausible deniability* about the choice among relational models in a range of situations where simple plausible deniability is absent. Note that the hypothesis is not that language is in fact context-independent or certain in transmission (the very existence of indirectness and vagueness shows that it is not), only that direct speech is *perceived* as such, and that this affects people's tuning of the directness of their speech.

Table 5
Payoff Matrix for Diner Without a Reservation, Including the Relational Models Assumed by Speaker and Hearer

Diner's strategy	Type of maitre d'	
	Dishonest maitre d'	Honest maitre d'
Don't bribe	Long wait (Authority/Authority)	Long wait (Authority/Authority)
Direct bribe	Instant seating (Equality/Equality)	Awkwardness (Equality/Authority)
Indirect bribe	Instant seating (Equality/Equality)	Long wait (Authority/Authority)

The first corollary of the digital-language hypothesis is that direct speech is perceived as *certain*. If, as Fiske proposed, relational models are discrete systems, then in the mutual signaling of which model applies, people must implement a mapping from a continuous domain of contexts to a discrete range of relational models. In other words, whereas the degree of certainty regarding the speaker's intention varies continuously, the prevailing relational model between two individuals can only be in one state or the other. How do two people come to a tacit agreement as to which model applies to their interactions?

This is an example of what game theorists call *coordination games*, scenarios in which several options are available to a pair of agents and they are both best off if they agree upon one of them, regardless of which it is (Clark, 1996; Schelling, 1960). In coordination games, agents often use *landmarks* or *focal points* as solutions. An example is the way that people negotiating over a price will often split the difference between their original positions or settle on a round number. In the case of people tacitly negotiating a relational model, the certainty of direct speech may similarly serve as a symbolic landmark, using the rest of the scale of indirectness as grounds for giving each other "the benefit of the doubt." For example, along the range of signals that probabilistically convey sexual intentions (physical proximity, suggestive remarks, relative seclusion, etc.), a woman may continue to treat the relationship as platonic all the way up to the point of an indubitable sexual proposition. Only at this point must she either rebuff the proposition, preserving the existing relationship type, or accept it. Under this corollary, indirect speech is useful as long as it allows for deniability to be *possible*, even when the deniability is not plausible.

A second corollary is that language is sensed to convey information *losslessly*. Just as a digital file containing text, music, or images can be transmitted and copied without degradation, the message conveyed by direct speech may be perceived as faithfully transmissible in chains of gossip—unlike indirect speech, where the underlying intention is uncertain and the context, history, and private details may be necessary to recover it (Clark & Schaefer, 1992). Gossip is widely recognized as an important component in the evolution of human cooperation (Nowak & Sigmund, 2005; Ohtsuki & Iwasa, 2006), and its prevalence in present-day social life supplies a motive for speakers to protect their reputations. Hearers in these encounters may also have similar motives, because communal disapproval of relationship-crossing transactions often extends to both parties. Thus, even a hearer known to the speaker with near certainty to be a willing cooperator may prefer that the request be tendered indirectly in order to maintain plausible deniability to outsiders or authorities. (Recall that in the dialogue from *Schindler's List* (Spielberg, 1993), the commandant hearer who granted Schindler's request tried to maintain plausible or possible deniability to third parties even in the speaker's presence.) Under this corollary, indirect speech is a means of protecting reputation from a virtual audience. Although the hearer may be nearly certain of the ulterior meaning and even act on it, the increasingly corrupted nature of the hearsay evidence available to third parties may attenuate any decrement in the reputations of the transacting parties. That is, with indirect speech, deniability is plausible *to a virtual audience*, even if it is plausible to neither the speaker nor the hearer.

A third form of higher-order plausible deniability arises from the fact that the syntactic rules of language define a recursive system: A

sentence may contain a sentence of the same kind (e.g., *She thinks that he knows that she likes him*; Chomsky, 1957; Pinker, 1994). This allows speakers to convey recursive propositions, in which one idea (*She likes him*) is embedded in another (*He knows that she likes him*). When two or more parties have recursive representations of one another's state of knowledge about some proposition, their understanding is called *mutual knowledge*, *common knowledge*, or *common ground* (Chwe, 2001; Clark & Marshall, 1981; Lewis, 1969; Smith, 1982).⁷ That is, two agents A and B have common knowledge of the proposition *x* when A knows *x*, B knows *x*, A knows that B knows *x*, B knows that A knows *x*, and so on. The reason that focal points, in particular, solve coordination problems is that they easily become common knowledge.

Logicians have shown that there are many logical differences between common knowledge and mere shared individual knowledge, but the simplest illustration comes from the story of the emperor's new clothes. When the little boy cried out that the emperor was naked, he did not tell the onlookers anything that they could not see with their own eyes. But he did change their state of knowledge nonetheless, because now everyone knew that everyone *else* knew that the emperor was naked, and this prompted them to challenge his authority through laughter. More precisely, if we denote the proposition THE EMPEROR IS NAKED by *x*, then we can see that the boy's exclamation augmented the presence of *x* in each hearer's knowledge base with the proposition *y*, by which we denote EVERYONE KNOWS *x* AND *y*, where *y* is recursively embedded in itself (Clark, 1996). For our purposes, the two morals of the parable are that direct language is an effective shared basis for common knowledge and that common knowledge is an effective way of proposing a change to a relational model. Common knowledge may even *inevitably* ratify such a proposal, overriding any intentions to the contrary. No matter how much fear the emperor's subjects may feel toward him—and hence no matter how strong the motive to pretend that they have not noticed the obvious—a state of common knowledge does not permit them this pretense without surrendering their tacit claims to rationality and honesty.

According to this corollary, then, direct speech (but not indirect speech) generates common knowledge, and relationships are maintained or nullified by common knowledge of intentions specific to the relational model. In this light, consider the man's invitation to view his etchings and his date's demurral. She may believe with near certainty that the invitation was a sexual solicitation, and he may believe just as strongly that she has turned it down. In other words, first-order or individual knowledge is present. But how certain is she that he understands the knowingness of her refusal? She may suspect that he is overly optimistic and cannot take a hint. On the other side, how certain is he that she has registered his understanding of her knowing refusal? "Maybe she thinks I'm dense," he might say to himself. As both speaker and hearer reel out successively deeper levels of the recursion, each level inherits the uncertainty from the previous levels and introduces some of its

⁷ Some writers use *mutual knowledge* interchangeably with *common knowledge*; others reserve it for shared individual knowledge and distinguish it from common knowledge. We use the unambiguous *common knowledge*.

own. This process can quickly lead to great uncertainty at a given level of common knowledge, even if confidence at the first level is high.

The creation of higher-order uncertainty may be an important objective of indirect speech. Although our moral psychology demands that we condemn clear violations of relational models, there are also costs associated with acknowledging such a violation, especially if speaker and hearer are enmeshed in circumstances that commit them to a continuing relationship. Moreover, it may be in the interest of both parties to avoid acknowledging the violation in order to keep alive the possibility of renegotiating the relationship in the future. These dynamics may help explain a phenomenon noted by Darwin (1872/1998): that blushing and other symptoms of awkwardness often cause the beholder to feel awkward as well. A *gauche* direct request can embarrass a hearer by placing her in the unwanted position of being an enforcer and constraining her future options. For these reasons people may be reluctant to acknowledge relational breaches lightly, and indirect speech can cater to this reluctance by attenuating the common knowledge that would render ignoring the breach an impossible charade.

Because the lack of common knowledge provides this “out,” unwilling hearers might raise their standard for how close an offer must be to a direct request before they are bound to give an appropriately antagonistic response. This may provide a sufficiently general reason for the higher standard of certainty required by antagonistic hearers before they must denounce the violation of the prevailing relational model, even in the absence of a tangible payoff structure that penalizes too-hasty whistle blowing.

These hypotheses about focal points and mutual knowledge build on Clark’s (1996) account of indirect speech, while differing from it in one crucial respect. Clark showed how an enormous body of common knowledge, including commonsense understanding of the physical and social world and command of the vocabulary and grammar of the language, is necessary for communication to proceed (he called this body of knowledge *common ground*). Each successful communicative act in the conversation adds to the common ground. In the handling of requests, a precondition for the commitment of both speaker and hearer is common knowledge of each party’s ability and willingness to carry out the joint project. An on-record prerequisite, designed to add the hearer’s ability or willingness to their common ground, can serve as an indirect proxy for the focal request itself. (For example, because a person can accede to a request to pass salt only if he or she is physically capable of passing salt, the prerequisite *Can you pass the salt?*, which adds an acknowledgment of that capability to common ground, can serve as a proxy for the imperative *Pass the salt.*) Acceptance of the prerequisite then locks both parties into an extended procedure to complete the joint project. But while this explanation applies naturally to on-record indirect speech, we propose that the crucial feature of an off-record indirect request is that it serves to keep the mental states of speaker and hearer *out of* common knowledge (common ground), granting both parties the freedom to treat the proposal and response in a personally convenient manner that may happen to contradict what the relational models proscribe.⁸

In sum, as a result of the qualitative nature of relational model choice and the perceived digital nature of language, indirect speech

allows for higher-order deniability of the challenge to a relational model. This higher-order deniability includes deniability that is possible (even if not plausible), deniability that is plausible to a virtual audience (even if not to the speaker and hearer), and deniability of common knowledge (even if not of individual knowledge). In this way both speaker and hearer maintain their face—which, in this context, encompasses the self’s projection of rationality, honesty, and proper regard for the relational models and other moral strictures.

Experiment 3

This experiment tests the strategic speaker theory as applied to relational model negotiation and higher-order deniability by assessing people’s judgments of what the characters in a fictitious scenario are thinking and feeling as they produce and interpret dialogue that varies in directness. The critical predictions are as follows.

First, the linguistic variable of directness (as assessed by raters in a pilot study) should correlate with participants’ assessments of the probability that a hearer interprets an indirect speech act as a relationship-changing request. (This is similar to the test in Experiment 2, but from the hearer’s point of view rather than the speaker’s.)

Second, the perceived probability that an antagonistic hearer would rebuff the request and act to the disadvantage of the speaker should also rise monotonically with directness, but at some levels the probability that an antagonistic hearer rebuffs the request must be lower than the probability that a cooperative hearer grants the request. (Recall that this inequality is necessary in the strategic speaker model for indirect speech to be optimal; it should also hold when the costs are not easily quantified in terms of dollar amounts and the like.)

Third, if speakers anticipate that hearers use certainty as a focal point in changing a relational model, and thus use indirect speech for possible (but not necessarily plausible) deniability, then participants should make a categorical distinction between direct speech and even highly suggestive indirect speech: The former should be judged as 100% certain in intent, the latter as less than 100%.

Fourth, if speakers anticipate that a relationship-changing proposition is less likely to be conveyed successfully by gossip if it is expressed by indirect than by direct speech (and thus use indirectness for plausible deniability to third parties), direct speech should be interpreted as confidently by a third party as by the hearer (indeed, with 100% certainty), whereas indirect speech should be interpreted with decreasing confidence with additional links in a gossip chain.

Fifth, if speakers anticipate that direct speech brings a relationship-changing proposition into common knowledge, and thus use indirect speech to achieve plausible deniability of that mutual knowledge, then participants should judge that direct speech is interpreted with the same certain or near-certain level of

⁸ A related speech act is the *ostensible invitation* (Isaacs & Clark, 1990), in which a speaker offers an insincere invitation, which the hearer recognizes as such and refuses, both tacitly understanding that the speaker has indicated he values the relationship. Isaacs and Clark (1990) noted that the invitation and refusal are “on record,” whereas the speaker’s motive is “off record.” These are what we would call common and individual (first-order) knowledge, respectively. We note that as with other off-record indirect speech acts, an ostensible invitation must be worded so as to leave uncertainty about the interpretation, with some probability that the listener will misinterpret it.

confidence regardless of the degree of embedding of the proposition in the mental states of the speaker and hearer.

Method

Participants. Thirty-one Harvard undergraduates filled out a paper questionnaire for course credit or pay (\$10).

Materials. The questionnaire presented the bribe, seduction, and threat scenarios used in Experiment 1, each one concluding with the speaker putting his request to the hearer.

The questionnaire contained four versions of all three scenarios, and each participant saw all four versions in a within-participants design. Each scenario text concluded with one of four phrasings of the request. In each case the four phrasings were chosen from an original list of eight on the basis of ratings solicited in the pilot study described in Experiment 2. The criteria for choosing the four phrasings in this study were the same: separation in a parallel boxplot and significant differences between adjacent sentences by the signed rank test. (Thus, the phrasings used in the bribe scenario were identical in Experiments 2 and 3.) The scenario text was followed by a series of questions eliciting participants' interpretations of the thoughts and feelings of the characters in the scenario.

The questions conformed to the schematic outline set forth below. (Additional questions pertaining to hypotheses not discussed in this article are omitted here but may be obtained by request.)

The speaker (S) makes a request to the hearer (H) through an indirect speech act. H declines. Later, H tells a third party (T) what S said.

1. Does H understand what S really means? [certainty as a focal point]
2. In this question and *all of the questions that follow*, assume that what S really meant by his last statement was indeed an illicit request, and that H ignores the request. Does S think that H understands what S really means? [common knowledge, 1st-order hearer]
3. How clear is S's real meaning in what he says? [reliability check]
7. Does H think that S thinks H understands what S really means? [common knowledge, 2nd-order hearer]
8. Suppose that S *does* realize that H knowingly turned down his request. Does S think that H thinks that S thinks that H understands what S really means? [common knowledge, 2nd-order speaker]
9. Suppose that H *is* certain that S knows H meant to turn down the request. Does H think that S thinks that H thinks that S thinks that H understands what S really means? [common knowledge, 3rd-order hearer]
10. Does T understand what S really meant? [virtual audience, 3rd-party]
11. Would S think that T understands what S really meant? [virtual audience, speaker's anticipation]

12. Miscellaneous scenario-specific question.

Question 3, a check on whether the participants agreed with our pilot sample's assessments of the relative directness of the speech acts, was answered on a Likert scale ranging from 1 to 7. For the other questions, participants circled one of seven options: 0%, 1%, 2%–49%, 50%, 51%–98%, 99%, 100%. We instructed participants to write down a specific percentage in a provided space if they choose either 2%–49% or 51%–98%, thus creating a quantitative scale of subjective certainty from 0 to 100%.

The actual wording of the scenarios was carefully designed to make the scenario plausible and to lead the participant one step at a time to the state of knowledge that is necessary to interpret the common knowledge questions. In particular, participants were asked to put themselves in the shoes of the speaker or hearer and answer in the first person, eliminating one level of remove from the characters' mental states. The component propositions were introduced sequentially, rather than in a single embedded sentence. Different verbs were used for each mental state of the speaker and hearer (e.g., *understand*, *think*, *know*, *realize*), rather than repeating the verb. And each number on the rating scale was explained with a full sentence that participants could agree or disagree with, rather than having them compose their own complex interpretation and map it onto the scale. The text of these questions in the bribe scenario is reproduced in Appendix B.

Procedure. Participants were run up to three at a time in a 1-hr session under the supervision of an experimenter. The order of the scenarios was randomized for each subject, whereas the order of the versions within each scenario was fixed. The versions of the scenario were never presented in order of increasing or decreasing directness of the culminating speech act. At the end of each session, participants were asked if they understood all of the questions. All answered in the affirmative.

Parallel boxplots were used to assess the effect of increasing directness on the outcome variables. Formal statistical testing was carried out by fitting linear mixed models to the responses as a function of directness level (treated as a quantitative variable with equal increments between levels), with participant variability as a random effect, by REML.

Results and Discussion

The responses were qualitatively similar across the three scenarios. In most cases the results for one scenario reveal all features of interest. Results of the formal statistical tests are reported in Table 6; all *p* values are less than .001. The table entries are regression coefficients that capture the effect size; they may be interpreted as the expected change in the participants' responses resulting from a one-step increase in the directness of the offer.

Plausible deniability of an offer that breaches a relational model. The strategic speaker theory predicts that the linguistic variable of directness correlates with the probability that a hearer will interpret the speech act in its ulterior sense. The responses to the first question, which asked participants to put themselves in the hearer's position and estimate the probability that the hearer would interpret the statement as a bribe, come-on, or threat, bear this out (see Figure 5).

The theory also predicts that these interpretations should be linked in the minds of the speaker and hearer with differing consequences.

Table 6
Estimated Effects of Directness on Interpretations of Propositions in Experiment 3

Question ^a	Hypothesis	Scenario		
		Seduction	Bribe	Threat
3		-1.40	-1.14	-1.74
12	Consequence of plausible deniability	-1.07 ^b	16.9 ^c	12.5 ^d
1	Certainty as a focal point	26.9	18.1	30.6
2	Common knowledge, 1st-order hearer	23.5	22.3	27.7
10	Virtual audience, 3rd party	21.8	14.9	27.5
11	Virtual audience, speaker anticipation	19.8	15.5	27.1
7	Common knowledge, 2nd-order hearer	21.3	17.0	25.6
8	Common knowledge, 2nd-order speaker	16.7	14.0	21.0
9	Common knowledge, 3rd-order hearer	18.6	12.7	20.8

^a The order of questions in the table mirrors the order in which the corresponding hypotheses were introduced in the text. ^b This question asked participants to rate on a 1–7 Likert scale the ease with which the man and his date are able to resume their previous nonromantic relationship. ^c This question asked participants to estimate the probability that the driver will be convicted at trial of attempted bribery. ^d This question asked participants to estimate the probability that the professor will be sanctioned by the university disciplinary board.

Question 12 in the bribe scenario asked subjects to give their estimated probability that a jury would find the driver guilty in a trial for attempted bribery. Question 12 in the threat scenario asked subjects to give their estimated probability that a university disciplinary board would sanction the professor for threatening the student. The responses are displayed in Figure 3. Participants clearly perceived, from their third-person vantage point, the advantage of phrasing a prescribed speech act in an indirect manner.

These are formal costs, like those in Experiment 2, and thus do not prove that the requisite cooperator–antagonist asymmetry in decision thresholds applies in purely informal social interactions governed by Fiske’s relational models. A modified version of the seduction scenario offers the possibility of a critical test. We presented the scenario to a new sample of 83 Mechanical Turk respondents and asked them to consider, separately, how Lisa would react to the various speech acts under two circumstances: (a) if she were attracted to Michael and would accept a proposition

to sleep with him, or (b) if she wanted to maintain a professional or platonic relationship, was offended by his proposition, and would rebuff this advance, shun him in the future, and gossip about him (see Appendix A). In two blocks of items, one for each possibility of Lisa’s values (order counterbalanced), participants rated the probability that she would react in the specified way for each of the four speech acts. As predicted, for each indirect speech act, participants thought that a willing Lisa’s acceptance of the proposition would be more probable than an unwilling Lisa’s open antagonism (see Figure 4; $p < .001$).

Certainty as a focal point and possible (vs. plausible) deniability. The effect of directness on the reported certainty of the hearer regarding the speaker’s intent was clearly positive in all three scenarios (see Table 6). The slopes of these best fitting straight lines, however, do not tell the whole story.

Figure 5 shows that all of the participants but one judged the direct overture in the bribe scenario to be interpreted by the hearer

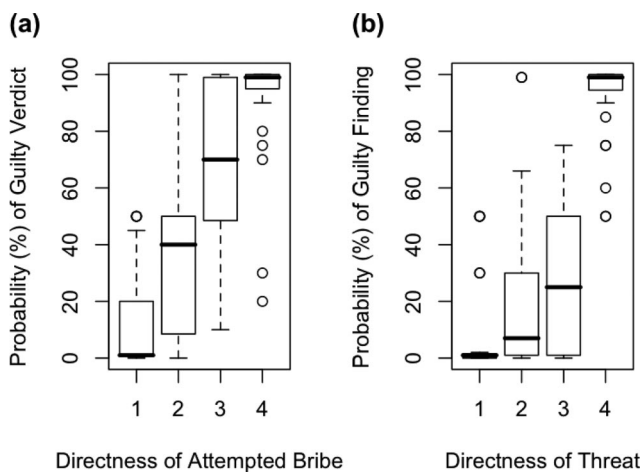


Figure 3. Parallel boxplots of Experiment 3 participants’ estimations of probabilities that (a) the driver will be convicted at trial of offering a bribe to the officer and (b) the professor will be sanctioned by a university disciplinary board for threatening the student.

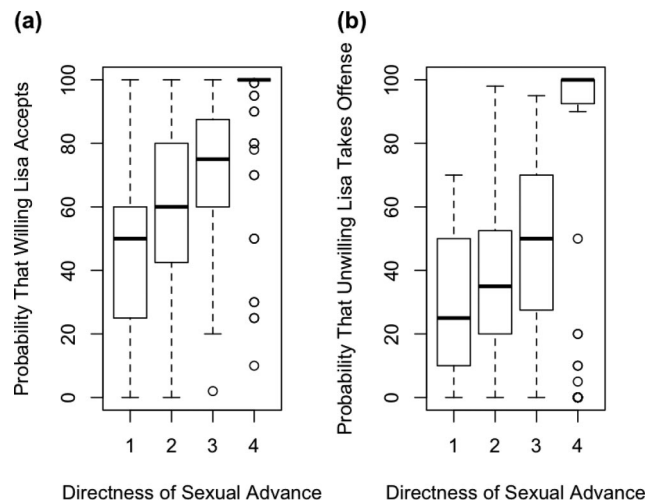


Figure 4. Parallel boxplots of participants’ estimations of probabilities that (a) a willing Lisa would accept Michael’s sexual proposition and (b) an unwilling Lisa would take offense.

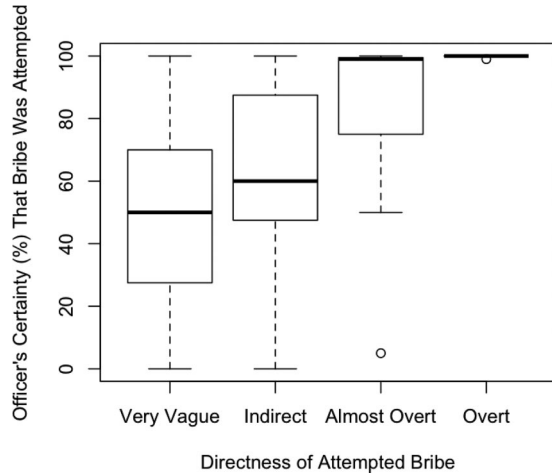


Figure 5. Boxplot of participant responses to Question 1 of the bribe scenario in Experiment 3: “How certain is the officer that the speech act is really an attempted bribe?” The speech acts are labeled explicitly rather than numbered in order of increasing directness.

with 100% certainty. The unanimity of the participants, and their choice of a single extreme value on the rating scale, is unusual in behavioral experiments and testifies to a strong perception of the digital nature of language—specifically, that a directly worded request is ubiquitously understood to indicate a speaker’s intention with certainty. Also unusual is the categorical differentiation between direct and indirect speech: Whereas the direct bribe was judged by all the participants but one as 100% certain, the thinly veiled bribe was judged by most of them as exactly one percentage point less certain: The mode, median, and 75th percentile of responses were at exactly 99%. This striking pattern suggests that there is a qualitative psychological difference between a direct proposition and even the most obvious indirect one, and it supports the hypothesis that one rationale for indirect speech is a kind of higher-order plausible deniability, in particular, the deniability of certainty. That is, the speaker uses indirect speech to skirt the focal point of certainty, thereby conveying a relationship-threatening request without forcing the parties to alter their assumed relationship.

As mentioned, the seduction scenario, with its absence of legal or financial penalties, offers several pure tests of the hypothesis that indirect speech is a cost-minimizing strategy in relational model negotiation. The experiment included an additional question for this scenario designed to test the effect of directness on the social relationship between speaker and hearer, in which participants were asked how easy it would be, in the wake of the request, for the man and woman to resume their normal friendship and day-to-day interaction. Figure 6 confirms that more indirect propositions were perceived as making it easier for them to do so. Although the medians of the two intermediate levels appear to contradict the overall trend, the mean difference between these two levels is not statistically significant.

The virtual audience and plausible deniability to third parties. Figure 7 shows the mean responses to the seduction scenario with respect to the extent to which the woman believes that the man intended to make a sexual advance and the extent to which the

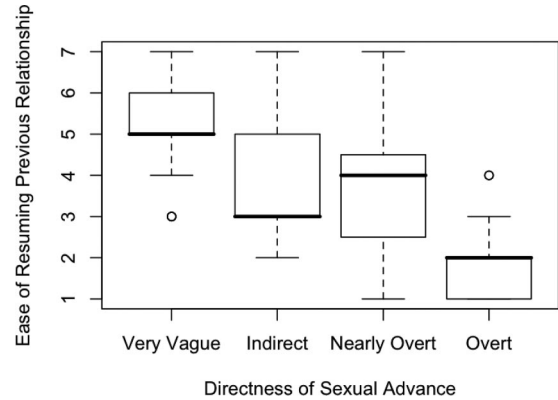


Figure 6. Boxplot of participant responses to Question 12 of the seduction scenario in Experiment 3: “How easily will the man and woman be able to resume their normal friendship and day-to-day interaction?”

woman’s friend believes the same after hearing about the encounter. Because the relevant consideration is not so much how a third party would interpret the speaker’s intent but how the speaker anticipates this interpretation, we also solicited judgments of the man’s anticipation of the friend’s interpretation. With direct speech, the level of certainty is pinned to 100% for the hearer’s interpretation, the third party’s interpretation, and the speaker’s anticipation of that interpretation. With the indirect speech acts, the certainty is less in all three cases. Note that with indirect speech, increasing distance from the actual time and place of the utterance results in increasing certainty: The woman’s friend was thought by participants to be more certain than the woman herself that the man intended something racy by his remark, despite the fact that she was not present, and the man thinks that the friend is more certain than she actually is. Similar trends are apparent in the other two scenarios. Several explanations of this anomaly are possible (perhaps the speaker is misled by immediate experience in a way that a neutral observer is not), but they are not essential to the test of the hypothesis that indirect speech affords

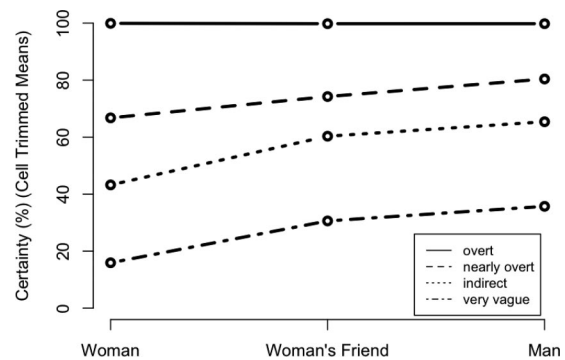


Figure 7. Line plots of participant responses in Experiment 3 (seduction scenario) to questions regarding the degree of certainty possessed by the hearer regarding the speaker’s intent, by a third party regarding the speaker’s intent, and by the speaker regarding the third party’s interpretation. Going from left to right, the tick marks on the x-axis correspond to Questions 1, 10, and 11. The top and bottom three responses within a cell were trimmed before calculating the cell’s mean. Increments along the x-axis are merely categorical.

plausible deniability to third parties. The relevant point is that with direct speech, no uncertainty exists in any direction: Present and absent parties are both completely certain of the intent, and the speaker knows it.

Deniability of mutual knowledge. To distinguish common knowledge from individual knowledge, it is necessary to separate the degree to which the speaker believes that the hearer believes something from the degree to which the hearer actually believes it (and so on for successive degrees of embedding of beliefs within beliefs). We thus phrased Questions 7–9 in such a way as to elicit the participants' judgments of the character's certainty at one level of the embedding, given certainty by the characters at all previous levels. For example, if *A* is the officer's probability (in the Bayesian sense, i.e., degree of subjective belief) that the driver's speech act was in fact an offer of a bribe, and if *B* is the officer's probability that the driver understands that the officer knowingly refused his offer, then the question regarding the latter proposition in effect asks for $P(B|A)$ —the officer's probability conditioning on absolute certainty that the driver's speech act was indeed an attempted bribe. Using the basic fact of probability that $P(B \cap A) = P(B|A)P(A)$, we can calculate a joint probability representing the participant's judgment of a character's degree of certainty at a given level of common knowledge after propagating the participant's assessments of uncertainty at all previous levels of the recursion.

For example, the officer's degree of certainty (as assessed by the participant) that the driver knows that he understood the speech act as an offer of a bribe depends both on the officer's belief in the driver's knowledge given that a bribe has indeed been offered and on the officer's belief that the driver was actually offering a bribe. Multiplying the participant's assessments of these two beliefs thus yields the participant's estimate of the degree to which the officer accepts the higher-order belief. Note that this procedure does not constrain the possible outcomes: Participants were free to indicate that a conditional belief is certain while the condition itself is doubtful, or vice versa, in any combination, and in different ways for direct and indirect propositions.

The effects of mental-state embedding on the interpretation of propositions varying in directness can be seen in Figure 8. The first

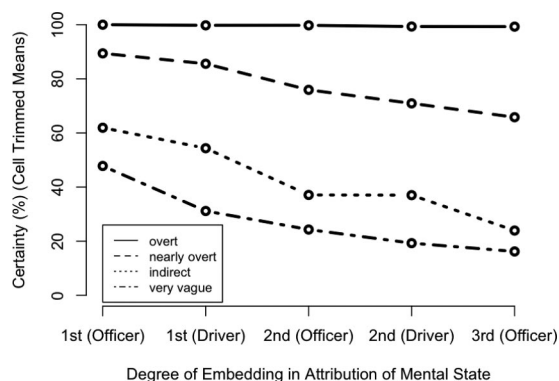


Figure 8. Judgments of the degree of first-order and common knowledge possessed by the driver and officer in Experiment 3. Going from left to right, the tick marks on the *x*-axis correspond to Questions 1, 2, 7, 8, and 9. For Questions 7, 8, and 9 the implied joint probabilities are shown. The top and bottom three responses within a cell were trimmed before calculating the cell's mean. Increments along the *x*-axis are merely categorical.

point in each line plot represents the officer's degree of certainty that the statement was an offer of a bribe, the second point the driver's degree of certainty that the officer knows that the statement was an offer of a bribe, and so on. Note that for the most direct speech act the estimates of perceived certainty remain essentially pinned to 100%, no matter the degree of embedding in the attribution of mental state. With all the indirect propositions, in contrast, the estimates of certainty decline with each level of recursive embedding, increasingly so as the statement becomes vaguer. The corresponding plots for the seduction and threat scenarios are similar. The perception that direct speech is certain at all levels of recursive embedding is all the more striking given the cognitive load imposed by embedded propositions. Though the difficulty of interpreting embedded degrees of belief may have reduced participants' degree of certainty that the speaker and hearer held those beliefs in the case of indirect speech, it clearly did not impede them when the same beliefs were conveyed by direct speech. This confirms the observation by Clark (1996) that people can apprehend a state of mutual knowledge directly, through the use of perceptual and linguistic cues, rather than explicitly spelling out the multiply nested propositions in their minds. Direct speech (but not indirect speech) appears to be one of those cues.

General Discussion

We have proposed a theory to explain why people so often insinuate a request indirectly rather than stating it baldly. The theory of the strategic speaker is based on the premise that off-record indirect speech is not just a social ritual but has a strategic rationale, and on the premise that language does not always involve pure cooperation between speaker and hearer but an uncertain mixture of cooperation and conflict. Off-record indirect speech is used to negotiate that uncertainty.

The core of the strategic speaker theory is a simple game-theoretic model showing that indirect propositions can allow for plausible deniability: A hearer favorably disposed to the request can accept it, whereas an uncooperative hearer cannot react antagonistically to it. The hypothesis that the use of indirect speech is a strategic response to the payoff structure defined by a mixture of cooperative and antagonistic hearers was supported in Experiment 2, which showed that speakers prefer to frame a bribe in more indirect forms as the hearer is more likely to be adversarial and a failure to strike a deal with a cooperative hearer becomes more costly. Similar support came from findings of Experiments 2 and 3, showing that the directness of a proposition is not just a linguistic variable governing the number of inferential steps that take one from the overt meaning to the implicated one but predicts the estimated probability that a hearer will interpret the proposition as expressing the implicated meaning. Crucially, both experiments found that antagonistic hearers required more directness than cooperative hearers before acting on their interpretations; this condition was a critical prediction of the hypothesis that indirect speech is an optimal strategy in identification games.

The theory was shown to apply not just to conversations with costs that are quantifiable in money or other tangible resources but to conversations that bear on the kind of relationship that holds between the speaker and the hearer (communal sharing, authority ranking, equality matching). The emotional costs of a breach in the current relational model can create a need for plausible deniability and thereby push a speaker toward indirectness even when there

are no overt costs. The informal observation that off-record indirect requests are ubiquitous in everyday life, despite the absence of quantifiable costs and rewards, was supported by judgments from Experiment 3 showing that in an imagined sexual proposition, indirect speech is perceived as more likely to allow the speaker and hearer to resume that relationship should a proposition be rebuffed, and that the critical asymmetry between antagonistic and cooperative hearers' threshold for action pertains to social as well as legally regulated actions.

Finally, the theory may be extended to higher forms of plausible deniability, owing to the discrete nature of relational models and the perceived digital nature of language. This combination allows a direct speech act to serve as a focal point in coordination games, to generate common knowledge, and to propagate a message with high fidelity along a chain of gossip. Indirect speech thus differs qualitatively from direct speech even when a speaker is highly confident of a hearer's values or a hearer is highly confident of a speaker's intent. The ability of indirect speech to avert the relationship-changing focal point of certainty was supported by a scatter of judgments among experimental participants regarding the probability that an indirect statement actually conveyed a request (always less than 100% certainty), in contrast with their near-unanimity that a direct one conveyed it with 100% certainty. Moreover, the ability of indirect speech to subvert relationship-changing common knowledge was supported by our finding a steadily deteriorating degree of certainty about the intent of an indirect request (but not of a direct request) as it was embedded in successively deeper layers of a speaker wondering whether a hearer knew what the speaker thought about what the hearer thought, and so on.

The theory of the strategic speaker builds on the insights of pure-cooperation theories such as P. Brown and Levinson's politeness theory and Clark's joint action theory, while differing from them in the treatment of off-record indirect speech. Both theories, we noted, do not deal explicitly with emotionally fraught and potentially conflictual verbal transactions such as bribes, threats, and sexual come-ons. And both conceptualize social transactions in dimensional terms, as negotiating a scalar quantity of face threat or equity imbalance. The theory of the strategic speaker, in contrast, proposes that indirect speech is a solution to potential conflicts of interest among speakers and that it negotiates a choice among qualitatively distinct and incommensurable relationship types. The theories were empirically contrasted in two ways.

According to politeness theory, indirectness is at the extreme end of a continuum of methods (including sympathetic politeness, deferential politeness, and conventionalized indirectness) by which a speaker redresses the threat to a hearer's face posed by a request. Experiment 1 showed that when the attitude of the hearer toward the request was identified within certain boundary conditions set by their relational model, increasing the face threat posed by a request was perceived to call for deferentially polite, but still direct, speech. Participants expected off-record indirectness only when a speaker addressed a hearer whose attitude toward the speech act was not identified and could range between the two extremes defined by different relational models.

According to joint action theory, speakers work to maximize their common acknowledgment of focal points and other forms of common knowledge (common ground). Experiment 3 showed that in the circumstances that call for off-record indirect speech, speakers work to avoid focal points and to keep certain knowledge *out of* common

ground. Indeed, keeping knowledge out of common ground is a transparent interpretation of the expression *off the record* itself: The metaphorical "record" is common knowledge or common ground.

The current results provide evidence of a new kind—from psycholinguistics—for the theory of Fiske (1991, 1992, 2000) that humans expect their relationships to fall into a discrete number of models and that a given relational model fosters resource transfers that promote the evolutionary rationale of the relationship and forbids those that contradict it, with self-conscious emotions as the enforcement mechanism. The finding that common knowledge and a virtual audience are important rationales for signaling these relationships underscores Fiske's observation that relational models are not agreed upon through negotiations between the parties but depend on a larger implicit understanding. That is, the rules of inclusion and exclusion implied by networks of communal sharing, the assignment of social rank, and the viability of reciprocal exchanges depend critically on the consensus of the community. Once the mechanisms enabling this consensus are implemented, a request in violation of a prevailing relational model may be greeted with indignant anger, in turn fostering preemptive embarrassment, awkwardness, or shame. However, because of the fluidity of human social dynamics, persons looking to reap the greatest possible gain from joint projects must test the bounds of their nominal relationship types and determine whether different ones might be consummated. In order to secure these provisional adjustments from willing cooperators while minimizing the possibility of being shamed by antagonists and gossipers, speakers resort to off-record indirect speech acts.

The harmony between the qualitative distinctions among relational models and the digital nature of language helps explain why a speaker can avoid exposure by an antagonist even in cases where the euphemism in which the request is couched seems flimsy and ineffectual. The difference in initial clarity between a barefaced request and a thinly veiled innuendo might be small, but the gap between them might lead the hearer to decide that a direct request is the only logical place to draw a threshold for declaring a violation of the prevailing relational model. The high fidelity of language also ensures that a bald proposition, but not an innuendo, retains its clarity as the speech acts reverberate through chains of third parties and recursive understanding. This becomes important if potential antagonists are unwilling to openly rebuff the party who first breaches their existing relational model because neither wants to follow through on the social and emotional consequences of acknowledging the breach. In the deliberately created absence of common knowledge, both speaker and hearer can test for a willingness to change the relationship or carry out minor expedient actions that contradict its rules, all without being forced to confront the face-effacing corollary that they are opportunistic, hypocritical, irrational, or easily exploited.

We suspect that the strategic use of indirect speech to keep certain problematic facts out of common knowledge can be extended to explain a variety of emotionally fraught social phenomena, including hypocrisy, taboo, tact, euphemism, piety, mock outrage, ostensible invitations, political correctness, and other examples of emperor's new clothes and elephants in the living room. To take one example, Bandura (1999) suggested that the use of euphemisms may enable people to perpetrate atrocities while protecting their commonly acknowledged identity as moral agents (see also Orwell, 1946). When hearing the dialogue from *Schindler's List* (Spielberg, 1993), one is struck by the euphemisms used to describe the slaughter taking place

inside the concentration camps: “special treatment,” “the processes that take place down here,” and so on. In one of the film’s final scenes, after the announcement of the German surrender, Schindler challenges the SS guards to carry out their orders to kill the thousand Jews assembled before them. “Or,” he says, “you could leave. And return to your families as men instead of murderers.” A long silence follows Schindler’s denuding of the emperor. Finally one of the guards breaks ranks and leaves. And then another. And then another.

References

- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3, 193–209.
- Brown, D. E. (1991). *Human universals*. New York, NY: McGraw-Hill.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. New York, NY: Cambridge University Press.
- Caplan, B. (2002). Systematically biased beliefs about economics: Robust evidence of judgmental anomalies from the Survey of Americans and Economists on the Economy. *Economic Journal*, 112, 433–458.
- Chomsky, N. (1957). *Syntactic structures*. The Hague, the Netherlands: Mouton.
- Chwe, M. S. Y. (2001). *Rational ritual: Culture, coordination, and common knowledge*. Princeton, NJ: Princeton University Press.
- Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). New York, NY: Cambridge University Press.
- Clark, H. H., & Schaefer, E. F. (1992). Dealing with overhearers. In H. H. Clark (Ed.), *Arenas of language use* (pp. 248–297). Chicago, IL: University of Chicago Press.
- Coen, E. (Producer), & Coen, J. (Director). (1996). *Fargo* [Motion picture]. Universal City, CA: Polygram Filmed Entertainment & Working Title Productions.
- Cole, P., & Morgan, J. L. (1975). *Syntax and semantics Vol. 3: Speech acts*. New York, NY: Academic Press.
- Darwin, C. (1998). *The expression of the emotions in man and animals* (3rd ed.). New York, NY: Oxford University Press. (Original work published 1872)
- Dawkins, R., & Krebs, J. R. (1978). Animals signals: Information or manipulation? In J. R. Krebs & N. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 282–309). Oxford, England: Blackwell.
- Dillard, J. P., Wilson, S. R., Tusing, K. J., & Kinney, T. A. (1997). Politeness judgments in personal relationships. *Journal of Language and Social Psychology*, 16, 297–325.
- Feiler, B. (2000, October). Pocketful of dough. *Gourmet*, 99–101.
- Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations*. New York, NY: Free Press.
- Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99, 689–723.
- Fiske, A. P. (2000). Complementarity theory: Why human social capacities evolved to require cultural complements. *Personality and Social Psychology Review*, 4, 76–94.
- Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology*, 18, 255–297.
- Gibbs, R. W. (1983). Do people always process the literal meanings of indirect requests? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 524–533.
- Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. New York, NY: Anchor Books.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics Vol. 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.
- Haidt, J. (2003). The moral emotions. In C. L. M. Keyes, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford, England: Oxford University Press.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour I and II. *Journal of Theoretical Biology*, 7, 1–52.
- Haslam, N. (1994a). Categories of social relationship. *Cognition*, 53, 59–90.
- Haslam, N. (1994b). Mental representation of social relationships: Dimensions, laws, or categories? *Journal of Personality and Social Psychology*, 67, 575–584.
- Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.
- Holtgraves, T. M. (1994). Communication in context: Effects of speaker status on the comprehension of indirect requests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1205–1218.
- Holtgraves, T. M. (2002). *Language as social action: Social psychology and language use*. Mahwah, NJ: Erlbaum.
- Holtgraves, T. M., & Yang, J. N. (1990). Politeness as universal: Cross-cultural perceptions of request strategies and inferences based on their use. *Journal of Personality and Social Psychology*, 59, 719–729.
- Horn, L. R. (2003). Implicature. In L. R. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 3–28). Malden, MA: Blackwell.
- Isaacs, E. A., & Clark, H. H. (1990). Ostensible invitations. *Language in Society*, 19, 493–509.
- Kasher, A. (1977). Foundations of philosophical pragmatics. In R. E. Butts & J. Hintikka (Eds.), *Basic problems in methodology and linguistics* (pp. 225–242). Dordrecht, the Netherlands: Reidel.
- Lakoff, R. (1973). The logic of politeness; or minding your P’s and Q’s. In C. Colum, T. C. Smith-Stark, & A. Weiser (Eds.), *Papers from the Ninth Regional Meeting of the Chicago Linguistics Society* (pp. 292–305). Chicago, IL: Chicago Linguistics Society.
- Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Lim, T. S., & Bowers, J. W. (1991). Facework: Solidarity, approbation, and tact. *Human Communication Research*, 17, 415–450.
- Maynard Smith, J. (1964, March 14). Group selection and kin selection. *Nature*, 201, 1145–1147.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, England: Cambridge University Press.
- Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, England: Oxford University Press.
- Nowak, M. A., & Sigmund, K. (2005, October 27). Evolution of indirect reciprocity. *Nature*, 437, 1291–1298.
- Ohtsuki, H., & Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239, 435–444.
- Orwell, G. (1946). Politics and the English language. *Horizon*, 13, 252–265.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York, NY: Morrow.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York, NY: Basic Books.
- Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York, NY: Penguin Books.
- Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences, USA*, 105, 833–838.
- Raggedclown. (2006, September 8). Dave Allen on the vagaries of the English language [Video file]. Retrieved from <http://www.youtube.com/watch?v=4If0UM6a4bA>
- Sampson, G. (1982). The economics of conversation. In N. Smith (Ed.), *Mutual knowledge* (pp. 200–210). Orlando, FL: Academic Press.
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Searle, J. R. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.),

- Syntax and semantics Vol. 3: Speech acts* (pp. 59–82). New York, NY: Academic Press.
- Smith, N. (1982). *Mutual knowledge*. Orlando, FL: Academic Press.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, England: Blackwell.
- Spielberg, S. (Producer & Director). (1993). *Schindler's list* [Motion picture]. Universal City, CA: Amblin Entertainment.
- Tannen, D. (1991). *You just don't understand: Men and women in conversation*. New York, NY: Ballantine Books.
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7, 320–324.
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78, 853–870.
- Tooby, J., & Cosmides, L. (1996). Friendship and the banker's paradox: Other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy*, 88, 119–143.
- Tracy, J. L., Robins, R. W., & Tangney, J. P. (2007). *The self-conscious emotions: Theory and research*. New York, NY: Guilford Press.
- Tracy, K. (1990). The many faces of face-work. In H. Giles & P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 209–226). London, England: Wiley.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.
- Trivers, R. L. (1985). *Social evolution*. Reading, MA: Cummings.
- Walster, E. H., Walster, G. W., & Berscheid, E. (1978). *Equity: Theory and research*. Rockledge, NJ: Allyn & Bacon.
- Wilson, M., & Daly, M. (1997). Relationship-specific social psychological adaptations. In G. Bock & G. Cardew (Eds.), *CIBA Foundation Symposium on characterizing psychological adaptations* (pp. 253–268). Chichester, England: Wiley.

Appendix A

Seduction, Bribe, Threat, and Favor Scenarios

Seduction

Michael and Lisa are coworkers and fairly good friends. Michael finds Lisa very attractive, but he has no idea whether she has any romantic feelings toward him. One day Michael asks Lisa if she wants to have dinner with him. Lisa agrees. Michael picks up Lisa at eight and drives her to a local restaurant. The food and wine are excellent, and the two of them have a great conversation. Michael picks up the check.

At ten-thirty Michael starts driving Lisa back to her apartment. He slows down while driving past his own apartment building and remarks that he lives only ten minutes away from her.

Bribe

Kyle is in a hurry to drive from San Francisco to Los Angeles. Since it is late at night and there are hardly any cars on Interstate 5, Kyle floors it and starts making great time. Before too long, however, Kyle is pulled over by a Highway Patrol officer. The officer comes up beside Kyle's window and shines a bright flashlight on Kyle's face. The officer says, "Hey, buddy. Did you know that the speed limit here is seventy miles per hour? You were doing over ninety." Kyle says, "I didn't realize I was going so fast." The officer says, "Well, you were. Please show me your driver's license."

Kyle has a history of moving violations, so he is worried that another ticket will boost his insurance costs and result in the suspension of his license. He slowly gets out his wallet and holds it out to the officer. The corner of a 50-dollar bill is protruding from his wallet ever so slightly.

Threat

Jennifer, a prospective graduate student, has submitted an application for an American Science Foundation Fellowship—an extremely prestigious honor with a financial award of over \$30,000 a year. Jennifer earned straight A's in college, scored a perfect 1600 on the GRE, and has already been a coauthor on

path-breaking articles published in the high-profile journals *Nature Genetics* and *PLoS Biology*. In short, Jennifer is a rising superstar and unquestionably worthy of a Fellowship.

One day Professor Jim Owens crosses paths with Jennifer in a hallway of the biology building. Jennifer was the best student in many of Professor Owens's courses, and they worked together on her senior thesis as true collaborators. This summer Jennifer has an offer to work in the lab of Josh Singer, which would provide her with the opportunity to broaden her experience and learn a new set of research skills.

Professor Owens asks Jennifer if she has a moment so that he can talk to her in his office. Jennifer says sure. Professor Owens happens to consult for the committee responsible for awarding Fellowships—something that Jennifer is well aware of.

Favor

[Note: In the version given here, the experimental factors of power gap, social distance, and degree of imposition are all set at their highest levels.]

Will has just started his high-pressure job as a financial analyst. One of his first tasks is to analyze data from the last few quarters and write up his findings in a report.

Will comes to work very early on the day that his report is due. At about 7 a.m. it dawns on him that a proper analysis of this time-series data requires expertise with stochastic processes. Will was enrolled in his college's course on stochastic processes for a single day; after staring blankly at the professor's mathematical scribbling for the entire class period, he promptly dropped the course.

Will steps outside his office to see who else might be at work this early. He finds that Brian is occupied with some task in his own office.

Will is a newcomer, so he barely knows Brian. Although Brian is not that much older than Will, Brian has a Ph.D in statistics and is currently the head of the company's research division. He was one of the company's three founders.

If Will does not finish his report in the next few hours, he will be in a world of trouble. The only way that he can finish the report is to ask Brian to go over the analysis together with him. Will knows that this would take around three hours. Brian would have

to clear his morning to help Will, even though the report would not benefit Brian's own division.

Will catches Brian on his way to the bathroom, gives him a friendly greeting, and quickly explains the situation.

Appendix B

Excerpt From Bribe Scenario in Experiment 3

Material in square brackets was not seen by the participants. Questions pertaining to hypotheses not discussed in this article are omitted. All questions were followed by 7-point Likert scales unless otherwise indicated.

[The text of the scenario as given in Appendix A.] Kyle says, [one of the following four:]

- "I'm very sorry, officer. I've really learned my lesson. From now on, you can be sure that I'll be more careful." [very vague]
- "I'm very sorry, officer. I know that I was speeding, and that I'll have to pay for my mistake." [indirect]
- "I'm very sorry, officer. But I'm actually in the middle of something right now, sort of an emergency. So maybe the best thing would be to take care of this here . . . without going to court or doing any paperwork." [nearly overt]
- "I'm very sorry, officer. If I give you a fifty, will you just let me go?" [overt]

[Instructions on how to answer the questions and fill out the rating scales.]

1. Put yourself in the officer's situation. Which of the following is the *most likely* thing that the officer is thinking at this point? [plausible deniability]

- 0%: The thought does not even cross the officer's mind that Kyle was offering him a bribe.
- 1%: "I'm virtually certain that this guy wasn't offering me a bribe."
- 2%–49%: "I think that this guy probably wasn't offering me a bribe." Indicate the officer's estimate of Kyle's degree of certainty here, using a number between 2% (probably not) and 49% (almost fifty-fifty):
- 50%: "Is this guy really asking me for help, shooting the breeze, or whatever? Or could it be that he's trying to bribe me? I really can't tell one way or the other."
- 51%–98%: "I think that this guy was probably offering me a bribe." Indicate the officer's estimate of Kyle's degree of certainty here, using a number between 51% (a tad more than fifty-fifty) and 98% (probably yes):
- 99%: "I'm virtually certain that this guy was offering me a bribe."
- 100%: "I'm absolutely certain that this guy was offering me a bribe (or at least as certain as anyone can ever be about someone else's intentions)."

2. In this question and *all of the questions that follow*, assume that (1) what Kyle really meant by his last statement was indeed an attempted bribe, and that (2) the officer ignores what Kyle said and continues the process of giving Kyle a ticket. Put yourself in Kyle's position. Which of the following is the *most likely* thing that Kyle is thinking at this point? [common knowledge, first order, speaker]

- 0%: "I'm absolutely certain that the cop didn't understand that I was offering a bribe (or at least as certain as anyone can ever be about someone's intentions)."
- 1%: "I'm virtually certain that the cop didn't understand that I was offering a bribe."
- 2%–49%: "I think that the cop probably didn't understand that I was offering a bribe." Indicate Kyle's estimate of the officer's degree of certainty here, using a number between 2% (probably not) and 49% (almost fifty-fifty):
- 50%: "Did the cop know that I was trying to bribe him? Or did he think that I was asking for help, shooting the breeze, or whatever? Hmm, it's a toss-up."
- 51%–98%: "I think that the cop probably understood that I was offering a bribe." Indicate Kyle's estimate of the officer's degree of certainty here, using a number between 51% (a tad more than fifty-fifty) and 98% (probably yes):
- 99%: "I'm virtually certain that the cop understood that I was offering a bribe."
- 100%: "I'm absolutely certain that the cop understood I was offering a bribe (or at least as certain as anyone can ever be about someone else's intentions)."

3. Did Kyle's spoken words convey his actual meaning clearly or ambiguously? [check on validity of directness manipulation]

7. Put yourself in the officer's situation. He knows that Kyle was really trying to bribe him, and he has turned down the offer. Which of the following is the *most likely* thing that the officer is thinking at this point? [common knowledge, second order, hearer]

- 0%: "This guy thinks that I didn't understand he was offering me a bribe. I'm absolutely certain of that (or at least as certain as anyone can ever be of another person's thoughts)."

• 1%: “This guy almost certainly thinks that I didn’t understand he was offering me a bribe.”

• 2%–49%: “This guy probably doesn’t think I understood that he was offering me a bribe.” Indicate the officer’s estimate of Kyle’s degree of certainty here, using a number between 2% (probably not) and 49% (almost fifty-fifty):

• 50%: “Does this guy know that I turned down his bribe attempt? Or does he think that maybe I didn’t get it, that I thought he was just making chit-chat? Hmm, I can’t really say.”

• 51%–98%: “This guy probably knows that I understood he was offering me a bribe.” Indicate the officer’s estimate of Kyle’s degree of certainty here, using a number between 51% (a tad more than fifty-fifty) and 98% (probably yes):

• 99%: “This guy almost certainly knows that I understood he was offering me a bribe.”

• 100%: “This guy knows that I understood he was offering me a bribe. I’m absolutely certain of that (or at least as certain as anyone can ever be of another person’s thoughts).”

8. Suppose that Kyle *does* realize that the officer knowingly turned down his attempted bribe. Put yourself in Kyle’s position. Which of the following is the *most likely* thing that Kyle is thinking at this point? [mutual knowledge, second order, speaker]

• 0%: “The cop thinks that I didn’t understand that he turned down my bribe. I’m absolutely certain of that (or at least as certain as anyone can ever be of another person’s thoughts).”

• 1%: “The cop almost certainly thinks that I didn’t understand that he turned down my bribe.”

• 2%–49%: “The cop probably thinks that I didn’t understand that he turned down my bribe.” Indicate Kyle’s estimate of the officer’s degree of certainty here, using a number between 2% (probably not) and 49% (almost fifty-fifty):

• 50%: “Does the cop know that I understood that he turned down my bribe? Or does he think that I didn’t understand? Really, it could go either way.”

• 51%–98%: “The cop probably knows that I understood that he turned down my bribe.” Indicate Kyle’s estimate of the officer’s degree of certainty here, using a number between 51% (a tad more than fifty-fifty) and 98% (probably yes):

• 99%: “The cop almost certainly knows that I understood that he turned down my bribe.”

• 100%: “The cop knows that I understood that he turned down my bribe. I’m absolutely certain of that (or at least as certain as anyone can ever be of another person’s thoughts).”

9. Suppose that the officer *is* certain that Kyle knows he meant to turn down the attempted bribe. Put yourself in the officer’s position. Which of the following is the *most likely* thing that the officer is thinking at this point? [mutual knowledge, third order, hearer]

• 0%: “This guy understands that I turned down his bribe. But he doesn’t realize that I know he understands that.”

• 1%: “This guy understands that I turned down his bribe. But he almost certainly doesn’t realize that I know he understands that.”

• 2%–49%: “This guy understands that I turned down his bribe.

But he probably doesn’t realize that I know he understands that.” Indicate the officer’s estimate of Kyle’s degree of certainty here, using a number between 2% (probably not) and 49% (almost fifty-fifty):

• 50%: “This guy understands that I turned down his bribe. But he may or may not realize that I know he understands that; I really can’t decide.”

• 51%–98%: “This guy understands that I turned down his bribe. And he probably realizes that I know he understands that.” Indicate the officer’s estimate of Kyle’s degree of certainty here, using a number between 51% (a tad more than fifty-fifty) and 98% (probably yes):

• 99%: “This guy understands that I turned down his bribe. And he almost certainly realizes that I know he understands that.”

• 100%: “This guy understands that I turned down his bribe. And he realizes that I know he understands that.”

10. Suppose that the officer tells another Highway Patrolman in his barracks the next day exactly what Kyle said, but doesn’t say what he thinks—he wants to know how his partner interprets what Kyle said. Put yourself in the position of the other officer listening to the first officer’s account. Which of the following is the *most likely* thing that the other officer is thinking at this point? [virtual audience, third party]

[Response options similar to those for Question 1 but stated in the first person by the other officer.]

11. Suppose that Kyle somehow learns that the officer chatted with his partner in this way (i.e., saying exactly what happened and what was said, but providing no interpretations or inferences). Kyle thinks to himself: “Hmm. If this other officer is certain that I offered a bribe, I guess he might go after me and get me in more trouble.” Put yourself in Kyle’s position. Which of the following is the *most likely* thing that Kyle is thinking at this point? [virtual audience, speaker’s anticipation]

[Response options similar to those for Question 10 but stated in the third person about the other officer.]

12. Suppose that the officer arrests Kyle for attempted bribery. Kyle’s trial is nine months later. In his testimony the officer tells the court and the jury exactly what happened and what was said during the night in question. That is all he testifies to; he does not give his interpretation or opinion about what anyone was thinking or what anyone really meant by something.

Kyle testifies in his own defense. His account of what happened is similar to the officer’s. He claims that the fifty-dollar bill was sticking out of his wallet on accident and that he did not mean to offer a bribe.

To find Kyle guilty of attempted bribery, the jury must find that the evidence establishes his guilt *beyond a reasonable doubt*. Which of the following do you think is the *most likely* outcome of Kyle’s trial? [consequences of plausible deniability]

• 0%: It is certain that the jury will acquit him (or at least as certain as any future event can be).

- *1%*: It is virtually certain that the jury will acquit him.
- *2%–49%*: The jury will probably acquit him. Indicate the probability that the jury will find him guilty, using a number between *2%* (probably not) and *49%* (almost fifty-fifty):
 - *50%*: The jury is just as likely to acquit him as to find him guilty.
 - *51%–98%*: The jury will probably find him guilty. Indicate the probability that the jury will find him guilty, using a number between *51%* (a tad more than fifty-fifty) and *98%* (probably yes):

- *99%*: It is virtually certain that the jury will find him guilty, but there is a very small chance that he will be acquitted.
- *100%*: It is certain that the jury will find him guilty (or at least as certain as any future event can be).

Received December 29, 2008
Revision received February 24, 2010
Accepted February 26, 2010 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!