Chapter 8
Self-Deceptive Speech: A Psycholinguistic View

Michael F. Schober and Peter J. Glick

In "On the psychology of self-deception," David Shapiro (1996) considers a man talking about a difficult decision he has made. The man says: "I *know* I did the right thing!" with exaggerated emphasis on "know," and he says it more loudly than he would in ordinary conversation. Shapiro characterizes this as the man's attempt to dispel his own doubts, as much to convince himself of the content of what he is saying as to convince his listener; it is thus an attempt at self-deception. In this case, it is an unsuccessful attempt, as evidenced by the man's following up, after a pause, with "I think."

Shapiro sees this as one example of self-deceptive speech, a kind of speech that has several qualities that distinguish it from speech with typical communicative aims. Self-deceptive speech of this sort can be characterized by repetition ("I *know* I did the right thing! I *know* I did!") and by surprising affective disconnects: either notably less affect in the saying than the content would typically warrant (e.g., speaking of being furious without any concomitant sign of anger), or alternatively what Shapiro describes as a melodramatic and artificial feel to the descriptions and gestures. As Shapiro (1996) recounts from prior clinical observation, e.g., from Hellmuth Kaiser in the 1950s, self-deceptive speakers do not "seem to express what they actually thought or felt. The tears sometimes seemed forced or worked up; the story of childhood sounded rehearsed; the angry account of yesterday's event, as one listened to it, had the quality of a public oration" (p. 788).

Another quality of this kind of speech, Shapiro (1996) proposes, is that it comes off as self-directed rather than listener-directed: the speaker is "addressing himself through the listener" (p. 790). Shapiro describes two manifestations of the inwardness of self-deceptive speech. In one version, the speaker does not seem to be attending to the listener at all. His speech has unusual prosodic characteristics that do not take the listener into account (as when our example speaker speaks unusually loudly), and he does not look at the listener in the ordinary way that speakers do. The listener can feel as if he or she is irrelevant to the speaker for the moment, and

M.F. Schober (✉)
Dean and Professor of Psychology, New School for Social Research, Department of Psychology, Office of the Dean, 6 E. 16th St., 10th floor, New York, NY 10003
e-mail: schober@newschool.edu

that the speaker would not notice any reaction the listener might have. In the second version, the speaker gazes intently at the listener as if hoping for confirmation and support for what he is trying to convince himself of; if that confirmation is not perceived as sufficiently forthcoming, through even slight hesitation by the listener, the speaker may work even harder at the self-convincing. Shapiro argues that despite the fact that in this second version the speaker is attending closely to the micro-reactions of the listener, the speech act is still self-directed rather than listener-directed, in that the listener is only being treated as a mirror for self-diagnosis rather than as a real interlocutor.

In Shapiro's view, self-deceptive speech is effortful, even if it is not consciously and deliberately planned, and it necessitates conscious activity because it involves speech and interaction. But speakers are not aware, or fully aware, of their self-deceptive activity; self-deceptive speech lies in a murky borderland between what is conscious and non-conscious, and what is effortful and automatic.

Shapiro's notion of self-deceptive speech does not find a straightforward counterpart in the mainstream of psycholinguistic views on the nature of speech planning and intention, nor in standard pragmatic models of communication. It is, by nature, a notion at the boundaries of ordinary language use. Even if the empirical basis of the phenomenon is not as firmly established as would be needed to convince corpus-based or experimental researchers of its generality, the issues it raises resonate in important ways with psycholinguistic questions and findings about the nature of and limits of language use. In this chapter, we explore these resonances, as well as describe our own empirical corpus-based explorations of self-deceptive language (or should it be called defensive language?), which demonstrate that it seems to have distinctive lexical characteristics (Glick and Schober 2007).

Finally, we will argue that under some circumstances the boundaries between self-deception, other-deception, motivated self-presentation that does not reach the level of deception, and simply choosing among legitimate alternate conceptualizations are quite unclear. Understanding when and how these differ will require greater clarity about the range and kinds of communicative situations in which speakers find themselves, as well as about how interlocutors contribute to speakers' self-deceptive utterances.

Self as Audience

Shapiro's notion of self-deceptive speech requires a speech system in which a speaker in the presence of an addressee can have herself as her primary audience, at least at the moment of the self-deception. This contrasts, presumably, with the ordinary communicative case where the speaker treats her addressee as a full interlocutor. At first blush this seems paradoxical: How can a speaker produce an utterance with the intention of conveying information to herself? Is it possible for a speaker to simultaneously know and not know what she intends? And is it possible *not* to consider one's partner in an interactive situation? Much current thinking about the nature of speech planning and production, and of interactive language

use, does not address these questions directly, but some proposals suggest what mechanisms or structures might allow such a bifurcation.

One issue at stake is the notion of a *communicative intention*. The standard approach (see Levelt 1989, for an excellent overview) is that speakers communicate intentions through speech, but not all intentions are communicated (as Levelt says, thank heaven! p. 59) and there are other ways of communicating intentions besides through speech. Communicative intentions, following Grice's approach, are special, in that they always involve an extra purpose of *intention recognition*. That is, for an intention to be communicative the speaker must also intend that the addressee understands what the speaker is saying *because of* the utterance – not because the addressee could otherwise infer the speaker's intention from other actions or behaviors or displays.

The intentions implicated in an utterance can be complex and multilayered. Beyond the main intentions in an utterance there can also be "side intentions" that are backgrounded as associations and embellishments (see Levelt 1989, p. 137). Communicative side intentions can include reasons for a speaker's actions, plans or decisions, and are often encoded in additional grammatical structures beyond the main ones (different clauses, different temporal frames). Non-communicative side intentions often involve the impression the speaker wishes to make on her interlocutor, most often positive social goals like being seen as knowledgeable, pleasant, powerful, humble, or competent. These kinds of self-presentational intentions are usually not encoded in the surface structure of a speech act; making them explicit and grammatical would undermine them. For example, a speaker desiring to be seen as pleasant would be unwise to make the direct claim of pleasantness.

What would this analysis make of the "I *know* I did the right thing!" example? On the surface, the speaker is making an assertion for the interlocutor with the communicative intention of asserting that he knows he did the right thing. If the speaker "really" believes this, then it is a straightforward communicative intention. If the speaker really does not believe this, then he is lying and this is a case of other-deception. He may also have side intentions for his interlocutor of appearing decisive and correct. The fact that Shapiro (and possibly the addressee present at the moment of the speaker's utterance) interprets the utterance as a failure shows that the side intention has gone awry; the inappropriateness of the paralinguistic form of the utterance (too loud) further helps undermine any such side intention. One could also propose that the very syntactic form of the utterance presupposes that there had been a prior accusation or question about the legitimacy of the speaker's judgment; for this (presumably invented) example we don't have a full record of prior discourse, but presumably this utterance would feel far less self-deceptive if it were a response to a direct question: "Do you know that you did the right thing?" Perhaps one part of what makes the utterance seem self-deceptive is that it is answering a question that hadn't been asked.

In any case, to consider this a case of self-deception, where the speaker is trying to convince himself of what he is saying, we must consider the speaker himself to be another addressee in the interaction, and to have a dual role of speaking and listening at the same time. Is there evidence for such a split?

Mainstream psycholinguistic theorizing actually *does* have a role for an internal addressee: what has been called the Monitor or Editor (see Levelt 1989, among many others). The idea is that in speech planning and execution there are monitoring

processes that keep track of the extent to which what one is articulating matches what one intends, at all levels of planning and execution (Levelt 1989): Does the ordering of what one is saying fit the larger communicative aims? Does the syntactic structure of one's message cohere? Does one's lexical selection fit what one is intending to say? Is one making any speech errors in grammar, word choice, or pronunciation? The evidence suggests that there are both internal self-monitoring processes that prevent certain kinds of errors from being produced overtly and external self-monitoring processes that catch errors after they have been uttered (and which then allow going back and correcting oneself).

Both internal and external self-monitoring processes require a kind of mental bifurcation, in that the monitoring allows a view of what is about to be or has been produced that is separate from the processes that produced it. In the case of external self-monitoring, the processes involved are most likely the same processes involved in comprehending speech uttered by another; that is, the speaker hears what she herself has said and can judge the extent to which what she is saying does or does not fit her intentions. (Evidence shows that the effectiveness of this kind of monitoring is reduced when the speaker's hearing is interfered with, see Lackner and Tuller 1979.) In the case of internal self-monitoring, the debates are trickier; some argue that a special set of monitoring processes need to be separate from other production and comprehension processes, while others would argue that not only does it make sense that the production system should use the same processes as the comprehension system, but that proposing separate systems leads to the kinds of logical homunculus problems that make dualistic theories of mind untenable (see Levelt 1989, for review). More recently, Pickering and Garrod (2004) have proposed that one of the main features of the mental processes of language users in dialogue is that speakers' production systems and listeners' comprehension systems are tightly interlinked, and that part of the interlinking lies in speakers' being able to attend to their own utterances in the same way that listeners do.

But for psycholinguists the role of the monitor is *not* typically assumed to act at the level Shapiro is talking about, where high-level intentions could be unknown to the speaker. Researchers in language production (see, e.g., Nozari and Dell 2009; Oppenheim and Dell 2008, among many others) have most often focused on "lower" levels of production–comprehension interaction as evidenced in phonological, lexical and syntactic errors and repairs. (It makes sense that this is where the research would have focused, as these kinds of issues are more tractable in empirical lab research and in analyses of large-scale corpora of speech errors.) The evidence from these kinds of studies is that the speech monitoring system is quite sensitive to the lexical characteristics of what is being said; speakers are more likely to produce speech errors that are words within the language rather than non-words, and more likely to produce errors that are semantically related to prior discourse and to the contextual situation (e.g., Baars et al. 1975; Motley 1980; Motley et al. 1981). Speakers seem not only to monitor for relatively "dry" aspects of the correctness of their speech output, but also for its social appropriateness. They are less likely to produce taboo words as speech errors than non-taboo words (Motley et al. 1982); the evidence suggests that the taboo word is actually created prior to articulation and the taboo word is then "retracted" before articulation in ways that produce particular speech errors.

In general, it looks as if there can indeed be relatively independent processes that produce an utterance and that then monitor or edit what is produced, either prior to articulation or afterward. (This point is not without controversy; monitoring functions could also be modeled as more central in the production process.) It also seems that the monitor can check for aspects of what has been produced that do not seem to be available to the initial processes that produced it; this suggests that there can be a split between what the producer "knows," at least temporarily, and what the monitor observes. But in all these cases the monitor is the wiser process, the one that comes along and cleans up the mess created by the less careful (or more time-pressured) production mechanism. It is not that the production mechanism is trying to convince the monitor of something that the monitor has yet to be convinced about.

So it seems that the internal audience for Shapiro's self-deceptive utterance is not the same sort of monitor that psycholinguists propose. In Shapiro's example, one could argue that the addition of "I think" is an after-the-fact effort reflecting the monitor's wisdom in recognizing how the initial utterance sounded, and so the monitor and the internal audience are the same thing. (And, simultaneously, the monitor is protecting the initial production mechanism from looking inappropriate or foolish to an external audience.) On the other hand, one could argue that, particularly in cases with no amendment like "I think," the Shapiro notion of self-deceptive speech is describing a failed monitor that does not recognize the discrepancy between what is intended and what is known, and does not see the utterance from any sort of external perspective. In any case, to our knowledge, there is no laboratory or corpus-based psycholinguistic evidence that demonstrates the kind of bifurcation in high-level intention production and comprehension that Shapiro proposes.

What the psycholinguistic literature *does* discuss are the kinds of failures to take one's audience into account that Shapiro argues go along with self-deception: speaking at inappropriate volume, saying more or less than the listener needs, failing to attend to one's partner. A sizable empirical literature has been investigating the extent to which and ways in which speakers succeed and fail at partner-directed language use (for reviews see Brennan and Hanna 2009; Brennan et al. 2010; Keysar et al. 1998; Krauss and Fussell 1996; Pickering and Garrod 2004; Schober 2006; Schober and Brennan 2003) in initial and subsequent moments of processing. The debates are far from resolved, but a notable proposal is that communicative language always starts out planned egocentrically and that speakers adapt to their partner's informational needs only at a later stage of the processing. (The debates are about when and whether this is so; mounting evidence suggests that when relevant information is available, speakers cognitively adapt to their partners from the very first moments of processing.) Following this line, one could characterize what is going on in the Shapiro examples as the speaker's failing to take the addressee's communicative needs and the social situation into account. That is, the addressee who feels that the speaker is not attending to them is picking up on a set of phenomena that do happen often enough in ordinary conversations. Under this interpretation the question would be how one should interpret the speaker's failure to adjust – is it characterological? Is it momentary? Or is it the norm, and only when all the circumstances are right do speakers fully adjust to their partners?

At a quite different level of analysis, one could conceive of the "I *know* I did the right thing" moment as a display that invites the interlocutor's support or confirmation, along the lines of the "ostensible" speech acts (see Isaacs and Clark 1990) that interactional psycholinguists consider. On this view, some speech acts, like many an invitation to "do lunch some time," involve a complicated multilayered social game, in which players who understand the game well know that what the speaker intends is not really what is on record (an invitation to lunch) but a kind gesture that ought to be turned down. Of course, the invitation is public and "on record" and so can't be withdrawn, which is why the game requires both parties to participate in the pretense if the game is to be played well, or else unintended results (lunch plans) or undesired offense (feeling rejected) can occur. On this sort of analysis, one could argue that the self-deceptive statement "I *know* I did the right thing," on record, is a statement of the speaker's belief; off record, it is a request for validation and affirmation that the speaker hopes the addressee will support. The "I think," as the kind of hedge that speakers can use to request input from their interlocutors (see Brennan and Ohaeri 1999), provides further evidence of the off record validation request. A willing partner in the ostensible act would collude by providing validation ("Of course you did the right thing!"). An unwilling or doubtful partner will not comply, and the speaker is stuck; he cannot, in any way that is socially acceptable, make explicit or insist upon the indirect validation request (just as the ostensible lunch-inviter cannot respond "Oh, I didn't really want to have lunch with you").

Obviously, we do not yet have a full picture of which psycholinguistic processes would be involved in self-deception as Shapiro describes it, either at the levels of internal and external self-monitoring or at the levels of partner adaptation or indirect speech acts. Although psycholinguists would describe interactive behaviors like inappropriate volume and failure to monitor the addressee's facial feedback as particular kinds of error in self-monitoring, there are not good accounts of *when* people succeed and fail at self-monitoring. There is good evidence that what speakers monitor varies during the course of an utterance, because it is not possible to monitor for everything all the time; there are cognitive limitations (see Levelt, chapter 12). And as a general principle when people are under greater cognitive load, or have greater working memory limitations, their processing becomes less efficient and accurate. So a reasonable proposal is that under notable cognitive load a speaker may fail to monitor. For Shapiro, the argument could be that trying to convince oneself of something that deep down you do not really believe leads to a particularly high cognitive load.

Lexical Features of Defensive Speech

Even without a full account of the psycholinguistic processes or structures that would be involved in truly self-deceptive speech, we can come closer to an account of what the qualities of that speech would be. As Shapiro has described the nature of self-deceptive speech, there are certain behavioral and paralinguistic features that go along with it, and there may also be structural features, like repetition. To

what extent have these clinical observations been validated using methods that psycholinguists would find definitive?

Although there is not much work of a psycholinguistic flavor (see Nelson and Horowitz 2001, for a rare example), there is clinical research on understanding the nature of defensive speech, which is a closely related (if not identical) notion (for discussion of whether all defensiveness is self-deceptive, and all self-deception defensive, see Barrett et al. 2002; Paulhus 1988; Sackeim 1988; Sackeim and Gur 1978; Shapiro 1989, among others). (We acknowledge that "defensive speech" has a more other-oriented flavor than "self-deceptive speech," and does not highlight the possible links between self-deception and other-deception that are worth exploring.) Building on this research, we (Glick and Schober 2007, 2011) have carried out analyses of the lexical characteristics of speech coded as defensive by clinicians, and we have found that it does seem to have some distinctive qualities compared to speech judged to be non-defensive.

In our study, we examined an audio corpus of 63 interviews of undergraduates that had been reliably coded for defensive behavior (Christensen 2003). In each interview, participants had been asked a series of 23 questions about potentially threatening experienced events, like "Describe a time when you've broken your own moral code." Later, researchers coded participants' audiotaped responses to these questions using a four-point scale that was designed to reveal defensive processes (Barrett et al. 2002), by focusing on the content, quality, and coherence of the verbal reports. Thus, we had a corpus of over 1,400 responses coded for how defensive they were.

We carried out lexical analyses on transcripts of the five questions in the interviews that elicited the most defensive answers using Pennebaker, Francis, and Booth's (2001) Linguistic Inquiry and Word Count (LIWC) tool, which is a dictionary-based word counting program. LIWC's dictionary, which was empirically developed and which has now been used in hundreds of studies, includes 74 different categories both of grammatical classifications (e.g., first person pronouns, past tense verbs) and content categories (e.g., family words, positive emotion words). We compared different responses to the same questions by people classified as more and less defensive, as well as different responses characterized as more and less defensive by the same participants; our analyses thus focused on 299 lengthy responses from the corpus.

The results indicated that participants classified as defensive used words differently than their counterparts classified as non-defensive. Across questions and participants, there was consistent and robust evidence that defensive speech had more words that LIWC classifies as cognitive mechanism words (such as *cause*, *know*, *ought*) and more exclusive words (such as *but*, *except*, *without*) than non-defensive speech. There was also some less conclusive evidence to suggest that defensive language was wordier, used more negation words (*no*, *never*, *not*), and repeated the same words more often.

Closer examination of what LIWC is picking up provides hints of what might be going on here. First, defensive speakers' greater use of cognitive mechanism words is consistent with the proposal that they may be attempting to protect their self-concept by offering reasons, justifications, and mitigations for their answers. For example, a

participant who had been asked about a time when she broke the rules offered the following explanation as part of her answer (cognitive mechanism words in italics):

> And so I *think* what they did was really unfair to me *because* usually what they do to students who who cheat is they they have them take the test all over again and in this case it was one little part on one little question and it was an equation. I *admit* it was wrong *but* the action they took was far too severe and so ever *since* then I've viewed the school as extremely hypocritical.

Here, the defensive participant refused to take full responsibility for her actions and instead externalized blame and culpability to an outside source. Such a justification seems to be indicative of defensive speech.

Second, in a similar fashion, the increased presence of exclusive words also may be attributed to defensive people's need to distance themselves from potentially incriminating or harmful information. In another example, a participant offered this as part of his response to describe a time when he broke his own moral code (exclusive words in italics):

> I don't like liars at all and it's not so much *that* I've lied recently *but* I've kind of hid things from my girlfriend now. It I mean it was for her own good I mean and she'll know today so *but* I've been like keeping it from her…

Here, the participant seems to defend himself against perceived criticism for lying to a significant other. To minimize the impact he reinterprets the events offering instead a modified view of his behavior. There is a partial admission of guilt but not without attempts at mitigating the severity of his actions.

Third, secondary findings from our study suggested that defensive speech may include other lexical markers. There was some evidence to suggest that defensive speakers were wordier and used more negation words, and that their utterances exhibited greater textual cohesion. Defensive justification may require more words to explain the reasons behind an action: a lengthy rationale to argue that the speaker's thoughts, feelings, or behaviors are not as detrimental or guilt-inducing as they sound. Negation words (*no*, *never*, *not*), like exclusive words, can be seen as evidence of a person's attempt to linguistically distance him/herself from an unwanted thought, feeling or behavior, and as such may aid in controlling the idea or perception of a disturbing external event (Vaillant 1992).

Of course, this is an analysis of defensive speech in a very particular setting – a research clinical interview – and so more will need to be done to find out whether these characteristics are similar to those found in other settings. And, of course, these analyses focus on the lexical characteristics of that speech, and not the paralinguistic or other aspects that Shapiro argues are important features of self-defensive speech. Nonetheless, they provide initial support for the idea that there are stable characteristics of a mode of discourse that is usually characterized more through case study analysis, and that with the right sampling methods we can start to understand those characteristics more systematically.

What is notable is that the lexical profile we see for defensive/self-deceptive language is different from that identified for other-deception (lying) (Berry et al. 1997; Burgoon et al. 2003; Hancock et al. 2008). When people lie, the evidence is that they

say less, use fewer first-person pronouns, more negative emotion words and fewer exclusive words (DePaulo et al. 2003; Newman et al. 2003). Unlike other-deceivers, our defensive speakers say more, use *more* exclusive words rather than less, and they do not differ in their use of first-person pronouns or negative emotion words from nondefensive speakers or when they are not speaking defensively. To the extent that our findings reflect a generalized profile of how defensive people talk, we can speculate that the processes involved in other-deception (which lead to the lexical form of what is said) are different from those involved in self-deception. That is, even though liars also seem to distance themselves from what they are saying, for example by referring to themselves less often, the kind of distancing they are doing may be different. This make sense; liars do not want the addressee to recognize that anything unusual is going on, whereas defensive speakers do, and they want the addressee's understanding or exoneration.

The research on other-deceptive language use has another intriguing finding that may be relevant to self-deception: the linguistic profile of the person being lied to looks different from the profile of a person who is being told the truth (Hancock et al. 2008). That is, there seems to be detectable evidence of the lie in what the interlocutor says, even when the interlocutor has no awareness that the speaker is lying. More specifically, the target of a lie tends to match their linguistic style to the liar more closely than when being told the truth. Some dimensions of this style matching include reducing their use of first person pronouns, causal terms and sentence complexity, while increasing their use of negations, just like the liar. The person being lied to also asks questions more when being lied to than when being told the truth, although this does not seem to help their ability to detect lies (Hancock et al. 2008).

The possibility that the recipient of self-deceptive or defensive language may actually show evidence of the self-deception in their own speech, whether or not they are aware that something unusual is going on, suggests that the interlocutor's role in defensive and self-deceptive language is worth considering more directly. We propose that a full account of self-deceptive speech should include its discourse context, and how the interlocutor's prior utterances, ongoing behavior, and reactions might be affecting the self-deceptive utterance.

What Kind of Communicative Moment?

The kind of speech that Shapiro examines falls into under-explored territory that is not discussed by the mainstream of theories of language use. We are curious about the range of situations in which this kind of speech happens, the kinds of interpersonal moments; surely there are moments in which it is less likely to occur. We assume, for example, that a speaker is less likely to say "I *know* I did the right thing" to a child at a birthday party or in a casual transaction at the supermarket with a clerk than in a confessional moment with a friend or spouse, an ethnographic interview, or a psychotherapy session. That is, even if the speaker is treating their addressee operationally as a sounding board rather than a real interlocutor, we

assume that there are nonetheless particular partners with whom the kinds of topics of discussion that could merit self-deception are still less likely to occur.

What our theories of language use lack is a taxonomy of situations that classify different kinds of interpersonal moments with different affordances, and thus which circumstances lend themselves to the expression of self-deception. To our thinking, it will be useful to understand more about those kinds of moments and their communicative dynamics, as it is not only the speaker's utterance but the addressee's being treated without regard in particular ways that characterize self-deceptive speech. Understanding when these moments shade into other-deception or are characterizable simply as "spin" or positive self-presentation will help flesh out a theory of self-deception.

Consider, for example, the moment during a clinical interview or survey interview when a patient or respondent is asked to report about a potentially sensitive or embarrassing behavior, as happens regularly when respondents are asked questions like: "How many sex partners have you had in the last 12 months?" The evidence is that different interviewing circumstances lead to different levels of reporting in large-scale surveys (and thus different national or regional estimates of sexual activity in the population): the average reported number of sex partners is notably different when people answer the question on a computer in a self-administered survey than when they answer to a human interviewer face to face or on the telephone; women report more sexual partners and men fewer (e.g., Tourangeau and Smith 1996; see Tourangeau and Yan 2007, for a review). Of course, we can't know which number is more likely to be the true answer, but the common interpretation is that most American women are embarrassed to report having had many sex partners and men are embarrassed to report fewer; somehow the self-administered computer interview frees respondents to reporting more accurately, perhaps because it is perceived as unable to judge or condemn. This sort of finding extends to various other kinds of sensitive questions and to questions about psychological distress, with respondents in some circumstances reporting more depressive and anxious symptoms to computers than to humans (see e.g., Epstein et al. 2001; Moum 1998; Rosen et al. 2009); and of course it raises questions about the usual wisdom that face to face interviews are the gold standard for creating rapport and eliciting honest responding.

Imagine that a respondent reports a particular number of sex partners to a human interviewer, and that she might have reported a higher number to a computer. What exactly might be going on here? There is a range of interpretations. One interpretation is that the respondent reporting the lower number is presenting a version of herself that is consistent with her desired self rather than her actual self, that is, that she is simply lying to look better. Another interpretation is that the respondent, faced with an intolerably embarrassing moment, is self-deceiving or lying to herself: telling herself what she wishes were true, even though a part of her knows that the truth is different than what she is presenting.

One could also account for this as neither deception nor self-deception, but rather as a legitimate (and truthful) self-presentation based on the respondent's leeway in deciding which behaviors to count as sexual acts. That is, while some physical

encounters unambiguously count as having a sex partner by any sensible definition, others are murkier and could be counted differently depending on one's personal or community definitions. Notoriously, there seems to be a generational difference on whether oral sex counts as sex, but various kinds of what was once called "petting" might be counted if one is particularly stringent or wants to report a high number of partners, but could be discounted if one wants to report a lower number. In this case, either a higher or lower number could be justifiable to report; either number could be consistent with providing an answer that is faithful to what the respondent believes the question author intended to be asking. Even if each answer represents a particular perspective or "spin" on the matter that presents the respondent in a different light, that is quite a different matter than other-deception or self-deception.

This kind of communicative issue arises more frequently than one might think, given the definitional variability across people and situations. For example, we have found (Suessbrick et al. 2005) that people's interpretations of what counts as smoking a cigarette are astonishingly variable when they are asked whether they have smoked 100 cigarettes in their entire lives; some count only tobacco cigarettes and others include clove cigarettes, cigars, and marijuana; some include any cigarettes they finished and others count cigarettes from which they took just one puff; some include only cigarettes they bought and others include cigarettes they "borrowed." The interpretation differences are enough to affect answers to the question; 10% of our respondents changed their answer from yes to no or no to yes when asked to count only tobacco cigarettes and count any from which they had taken even a single puff. In this case, we would be hard pressed to consider these alternate reports as deceptive or self-deceptive, or even unsavory "spin"; they seem to reflect different boundaries for what counts in a category and what does not, and perhaps different levels of flexibility in how solid those boundaries are in different communicative situations.

In contrast, a different moment in our tobacco interviews seems, on the surface, far more connected with a sense of self-deception. Current smokers in the survey were all asked whether they intended to quit smoking within the next 12 months. All of them answered that yes, they did. Next they were asked whether they intended to quit smoking within the next 6 months. A large majority reported that they did not, although often after a pause, and quite often with a rueful laugh, as if they had been caught at having been insincere a moment earlier when reporting their intention to quit in 12 months. That moment of saying "yes" to intending to quit in 12 months seems a legitimate candidate as a moment of self-deception given the immediately subsequent answer and the affect that went with it. But one could also ask whether it should better be considered other-deception, or politeness (wanting to be agreeable to the hint that the question assumes) or motivated self-presentation: wanting to present oneself in the best light to a human interlocutor, or wanting to be the person one wishes to be. Or, more complicatedly, whether in the murky moment of answering "yes" several of these might be at play.

In any case, we propose that considering a range of examples of polite responses, clear lies, self-presentations with "spin," and cases of speech characterizable as defensive or self-deceptive will help us to lay out this particular interpersonal

terrain, and clarify which moments are unambiguous and which fall into the communicative borderland. Shapiro's characterization of self-deceptive speech presents a challenge to the mainstream of psycholinguistic accounts that is worth taking up. We propose that understanding more about the audience for these acts, what the speaker believes about the audience, and how the audience's ongoing behavior affects the speaker are likely to be an important part of the puzzle.

References

Baars, B. J., Motley, M. T., & McKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *14*, 382–391.

Barrett, L. F., Williams, N. L., & Fong, G. T. (2002). Defensive verbal behavior assessment. *Personality and Social Psychology Bulletin*, *28*, 776–788.

Berry, D. S., Pennebaker, J. W., Mueller, J. S., & Hiller, W. S. (1997). Linguistic bases of social perception. *Personality and Social Psychology Bulletin*, *23*, 526–537.

Brennan, S.E., Galati, A., & Kuhlen, A. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation*, *Vol. 53* (pp. 301–344). Burlington: Academic Press.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274–291.

Brennan, S. E., & Ohaeri, J. O. (1999). Why do electronic conversations seem less polite? The costs and benefits of hedging. *Proceedings of the International Joint Conference on Work Activities, Coordination, and Collaboration (WACC '99)* (pp. 227–235). San Francisco: ACM.

Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F. (2003). Detecting deception through linguistic analysis. *Intelligence and Security Informatics*, *2665*, 91–101.

Christensen, T. C. (2003). *Experiencing sampling procedures: Are they probes to autonetic awareness?* Unpublished doctoral dissertation, Boston College, Boston.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.

Epstein, J. F., Barker, P. R., & Kroutil, L. A. (2001). Mode effects in self-reported mental health data. *Public Opinion Quarterly*, *65*, 529–549.

Glick, P. J., & Schober, M. F. (2007). *Lexical content of defensive speech*. Paper presented at the 17th annual meeting of the Society for Text and Discourse, Glasgow.

Glick, P. J., & Schober, M. F. (2011). *Lexical characteristics of defensive speech*. Manuscript in preparation.

Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M. T. (2008). On lying and being lied to: A linguistic analysis of deception. *Discourse Processes*, *45*, 1–23.

Isaacs, E. A., & Clark, H. H. (1990). Ostensible invitations. *Language in Society*, *19*, 493–509.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*, 46–50.

Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 655–701). New York: Guilford.

Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale: Lawrence Erlbaum.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.

Motley, M. T. (1980). Verification of "Freudian slips" and semantic prearticulatory editing via laboratory-induced spoonerisms. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic.

Motley, M. T., Baars, B. J., & Camden, C. T. (1981). Syntactic criteria in prearticulatory editing: Evidence from laboratory-induced slips of the tongue. *Journal of Psycholinguistic Research*, *10*, 503–522.

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*, 578–594.

Moum, T. (1998). Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression. *Social Indicators Research*, *45*, 279–318.

Nelson, K. L., & Horowitz, L. M. (2001). Narrative structure in recounted sad memories. *Discourse Processes*, *31*, 307–324.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin*, *29*, 665–675.

Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a "lexical editor" be? *Journal of Memory and Language*, *60*, 291–307.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*, 528–537.

Paulhus, D. L. (1988). Self-deception: Where do we stand? In J. S. Lockard & D. L. Paulhus (Eds.), *Self-deception: An adaptive mechanism?* (pp. 251–257). Englewood Cliffs: Prentice Hall.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah: Erlbaum.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–190.

Rosen, R. L., Schober, M. F., & Conrad, F. G. (2009). *Mode effects in questions about stigmatized behaviors and personal distress*. Paper presented at the 64th annual conference of the American Association for Public Opinion Research, Hollywood.

Sackeim, H. A. (1988). Self-deception: A synthesis. In J. S. Lockard & D. L. Paulhus (Eds.), *Self-deception: An adaptive mechanism?* (pp. 146–165). New York: Prentice-Hall.

Sackeim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. 2, pp. 139–197). New York: Plenum.

Schober, M. F. (2006). Dialogue and interaction. In K. Brown (Ed.), *The encyclopedia of language and linguistics* (2nd ed., pp. 564–571). Oxford: Elsevier.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Mahwah: Lawrence Erlbaum.

Shapiro, D. (1989). *Psychotherapy of neurotic character*. New York: Basic Books.

Shapiro, D. (1996). On the psychology of self-deception. *Social Research*, *63*, 785–800.

Shapiro, D. (1996). On the psychology of self-deception. *Social Research*, *63*(3), 785–800.

Suessbrick, A., Schober, M. F., & Conrad, F. G. (2005). When do respondent misconceptions lead to survey response error? In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria: American Statistical Association.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format and question context. *Public Opinion Quarterly*, *60*, 274–304.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*, 859–883.

Vaillant, G. E. (1992). *Ego mechanisms of defense*. Washington, DC: American Psychiatric Press.

Reply to Michael Schober and Peter Glick

David Shapiro

There can hardly be a subject matter more important to effective clinical work than self-deceptive speech. The reason is simple, though it involves a major change in the conception of the dynamics of psychopathology. We are able to see now that those dynamics are not constituted of particular unconscious conflicts within the person, separated from that person's attitudes and purposeful behavior. To the contrary, the individual's purposeful action is involved in a central, self-protective way in the pathological dynamics. It is often rationalized and in any case is aimed unwittingly at dispelling anxiety whose nature is unknown to its subject. This self-protective action is represented in the therapy hour itself and consists largely of self-deceptive speech. In that speech the patient reassures himself, without realizing that he is doing so; defends himself against charges, not knowing that they are of his making; persuades himself that he really wants what he thinks he should want. Well trained therapists are able recognize these dynamics, at least some of the time. But, how? We do not know all that makes self-deceptive speech recognizable or what, if anything, makes it distinguishable from conscious deception. This is the area that Schober and Glick look into and in which they offer an experimental study. It is clearly an area where dynamic psychology and psycholinguistics meet.

As Schober and Glick point out early in their essay, the psychology of self-deception requires some conception of self-monitoring of speech and thought, and therefore, as they put it, some kind of "mental bifurcation" is implied. The complicated question arises of whether the internal monitor is to be imagined as a separate institution. Schober and Glick, if I understand them correctly, lean carefully toward the assumption of independent processes for the production of speech and monitoring it, although they are obviously conscious of the logical risks of a dualistic conception. Clinicians, however, have a special reason for rejecting the idea of a separate self-monitoring institution. Homunculus-like conceptions of internal forces or agencies to which responsibility for action is assigned have played a problematic role in dynamic psychology. I am referring in particular to the conception, especially present in early psychoanalysis, of unconscious forces or agencies pursuing *their* aims in symptomatic behavior. It is a conception that works directly against the therapeutic goal of reviving the patient's experience of *his* reasons, not those of an unconscious force within him, for doing what he does.

The concept of a separate self-monitoring agency for self-deception, an agency that knows what not to know, is not only logically dubious, but also unnecessary. As I said in my paper *On the Psychology of Self Dece*ption (1996), "Regulatory monitoring and even regulatory action do not necessarily require understanding, and intention does not have to be knowledgeable. We jerk our hand from the hot plate not because we are afraid of damage to the skin, but because it hurts" (p. 786).

In other words, an internal monitoring effect requires only a signaling system, not a knowledgeable one. Specifically, it requires only an organization of attitudes which will respond with discomfort to the incipient articulation of ideas inimical to it. The individual's character or personality constitutes such a monitor. Inasmuch as such a discomfort or anxiety will then trigger, according to the personality, a corrective reaction, such as self-deceptive speech ("I *know* I did the right thing"), capable in principle of forestalling that conscious articulation, the monitoring and corrective actions are one. (Schober and Glick, incidentally, presume this example of mine to have been invented. It was not. The speaker was a free lance professional who frequently faced difficult job choices. In this instance, he had rejected the bird-in-the-hand at least partly because he considered that choice to be "weak.")

Schober and Glick consider the monitoring function in self-deception to be of a different sort from the kind familiar in psycholinguistics for the avoidance of speech errors. Their reason for that view is not clear to me. I should think that a signaling conception of monitoring comparable to the process I proposed for self-deception would be applicable as well for the correction of language errors. That would require, presumably, an existing structure of lexical, and perhaps social, standards sensitive to errors or transgressions. While the subjects and the effects of monitoring are certainly quite different in the two cases, their formal relationship may not be so distant. I say that admittedly prompted by a general supposition that the processes involved in self-deceptive speech must be ordinary psycholinguistic processes put to special use.

It is a curious thing about the experimental study Schober and Glick report that from the standpoint of a clinician it was more successful than might have been expected. To a clinical observer, after all, the most obvious distinction between self-deceptive speech and ordinary communicative speech (and lying) is not in the difference of the words used, but the different ways the same words are used. For that reason, I would think that formal or structural features of speech are more likely to show distinctions between self-deceptive and communicative speech than the use of particular words. Schober and Glick mention that possibility also and demonstrate it with findings such as greater wordiness and repetition in self-deceptive speech. Perhaps the use of many qualifying clauses would be another such feature. Of course, the separation between particular word content and structural features is not absolute; certain words like "but," "really" or "never," possibly the unusual use of adverbs and adjectives in general, suggest unnecessary emphasis or wordy complication. At any rate, though they may have fished in unpromising waters, Schober and Glick have made a catch. They have produced evidence of lexical distinctions and in that way have confirmed, as they say, the distinctness of a self-deceptive style of speech.

Schober and Glick raise additional interesting questions concerning the effects of the particular audience on self-deceptive speech and vice versa. Altogether, the relationships of the self-deceiving speaker to himself and to his listener are of great interest. In this connection, I found especially interesting their anecdote telling of the smoker's rueful laugh during the survey when he recognized, and abandoned, his self-deception. I have seen a good natured and rueful laugh of exactly that kind many times as a psychotherapist at the moment a patient becomes conscious of his

own self-deception. It is the patient's laugh at himself as he abandons an effort that he had not noticed himself making and now, seeing it, regards as pointless. That laugh in psychotherapy – and I have no doubt also in the survey interview – expresses a significant alteration not only in the individual's emotional state of mind but, also, in his cognitive state. His voice sounds more relaxed and conversational. He no longer looks away as he speaks or looks only for signs of confirmation; he looks with recognition at the listener and the listener in turn has the distinct impression of now being seen. This cognitive change occurring as it does at the moment the self-deception is abandoned brings into relief the speaker's peculiar cognitive state before its abandonment.

Reply to David Shapiro

Michael F. Schober and Peter J. Glick

We appreciate David Shapiro's thoughtful reply to our chapter, and we agree with him that any processes involved in self-deception should indeed be extensions of basic cognitive and psycholinguistic processes. Like him (and various psycholinguists), we are also uncomfortable with an overly dualistic notion of the monitor. Nonetheless, we believe that it will be a challenge for a theory of self-deception to detail how a monitor for high-level intentions really can be construed in a non-dualistic way. Even though it would be parsimonious if all monitoring processes, from lowest to highest levels, were continuous in terms of mental representation and processes, there is as yet no evidence that demonstrates such continuity. How one conceives of the "self" in self-deception that is deceivable is thus a complicated affair. We believe that Shapiro's alternative "reflex" notion (of a non-knowledgeable signaling system) is attractive, but many more details are left to be worked out before such a model fully connects with what is known about language planning and processing.

As for what Shapiro calls our fishing expedition, we are less surprised than he that we find in our data lexical evidence for a defensive style of speech, as lexical markers are concomitants of syntactic and discourse-level phenomena. Nothing in our findings requires that this style of speech be lexically driven rather than being a "linguistic fingerprint" or trace of the cognitive phenomena involved. On the other hand, language production involves many intertwined layers or cascades of macro- and micro-planning, with discourse-level intentions and syntactic choices not only driving but being driven by lexical choices. So we would not rule out that these effects could be lexical or at least have lexical components; we would expect to see evidence for a discourse style at multiple levels.

We are glad to know that the example "I *know* I did the right thing!" wasn't invented. From our perspective, it would be fascinating to see a full transcript not only of the patient's but also of Shapiro's verbal and paralinguistic behaviors throughout the interaction; it would also be useful to understand more of what happened before. We assume that there could be much to learn from a close look at what the therapist contributes to such a moment – both in the patient's beliefs about how the therapist is likely to respond (which could be based as much on carry-over from other conversational partners as from the therapist) and in how the therapist's behavior at the moment could make a difference. Presumably the effects of the therapist's overtly challenging the self-deceptive speech and colluding in it would be evident; presumably if the therapist merely listened with no visible change in expression this could also be interpreted as a signal by the patient. Even the physical setting – chair or couch – could presumably make a difference not only in what the patient thinks is going on, but in how responsive the therapist is and how the patient

interprets a therapist's silence (DiNardo et al. 2005). As we see it, the very particular conversational goals that the interlocutors bring to this conversation – which may not match (Russell and Schober 1999) – surely are at play in making sense of the peculiar kind of moment into which Shapiro so intriguingly delves.

References

DiNardo, A. C., Schober, M. F., & Stuart, J. (2005). Chair and couch discourse: A study of visual copresence in psychoanalysis. *Discourse Processes*, *40*, 209–238.
Russell, A. W., & Schober, M. F. (1999). How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes*, *27*(1), 1–33.