Synthese (2010) 176:399–427
DOI 10.1007/s11229-009-9573-1

In defence of gullibility: the epistemology of testimony and the psychology of deception detection

Kourken Michaelian

Received: 3 October 2008 / Accepted: 12 May 2009 / Published online: 24 May 2009

Abstract Research in the psychology of deception detection implies that Fricker, in making her case for reductionism in the epistemology of testimony, overestimates both the epistemic demerits of the antireductionist policy of trusting speakers blindly and the epistemic merits of the reductionist policy of monitoring speakers for trustworthiness: folk psychological prejudices to the contrary notwithstanding, it turns out that monitoring is on a par (in terms both of the reliability of the process and of the sensitivity of the beliefs that it produces) with blind trust. The consequence is that while (a version of) Fricker's argument for the necessity of a reduction succeeds, her argument for the availability of reductions fails. This does not, however, condemn us to endorse standard pessimistic reductionism, according to which there is no testimonial knowledge, for recent research concerning the methods used by subjects to discover deception in non-laboratory settings suggests that only a more moderate form of pessimism is in order.

Keywords Epistemology · Testimony · Monitoring · Deception

1 Monitoring and blind trust in the epistemology of testimony

In a pair of oft-cited papers (Fricker 1994, 1995)—see also Fricker (1987, 2002, 2004, 2006a,b,c)—Fricker makes an intuitively compelling case for local reductionism in the epistemology of testimony. Against the antireductionist, she argues that a reduction of testimonial justification[1] (the justification of testimonial beliefs) is necessary: if her

[1] Or testimonial knowledge; see Sect. 2. I will when no confusion will result speak for the sake of simplicity in terms of justification only.

K. Michaelian (✉)
Philosophy Department, University of Massachusetts, Amherst, MA 01003-9269, USA
e-mail: michaelian@philos.umass.edu

testimonial belief (where a testimonial belief is one which results simply from understanding and accepting a testimonial utterance) is to be justified, a subject must not, when she receives the relevant testimonial utterance, simply assume, as the antireductionist says that she may, that the speaker is trustworthy; to make that assumption, to trust the speaker blindly, is to believe gullibly. Against the global reductionist, Fricker argues that only a local reduction of testimonial justification is necessary: the subject need not establish the trustworthiness of testimony in general but only to establish the trustworthiness of the relevant speaker on the relevant occasion and with respect to the relevant topic; to look for a global reduction of testimonial justification is needlessly (indeed hopelessly) ambitious. Fricker argues, finally, that local reductions of testimonial justification are possible: a subject who monitors speakers for signs of untrustworthiness does not simply assume that a speaker is trustworthy but rather believes with empirical reason that she is so; thus we need not posit a right to assume trustworthiness in order to account for testimonial justification.

I grant that given that a reduction of testimonial justification is wanted, we need look only for a local reduction, and I thus have nothing to say against the second of these arguments; my focus here is rather on the first and the third of the arguments. Research in the psychology of deception detection—the term is used in the literature as a synonym of 'lying' or 'dishonesty'—implies that Fricker overestimates both the epistemic demerits of the antireductionist policy of trusting speakers blindly and the epistemic merits of the local reductionist policy of monitoring speakers for trustworthiness: folk psychological prejudices to the contrary notwithstanding, it turns out that monitoring is on a par (in terms both of the reliability of the process and of the sensitivity of the beliefs that it produces) with blind trust; there is much more to be said in defence of "gullibility" (and against monitoring) than Fricker supposes. The consequence is that while (a version of) Fricker's argument for the necessity of a reduction succeeds, her argument for the availability of reductions fails. This does not, however, condemn us to endorse standard pessimistic reductionism, according to which there is no testimonial knowledge, for recent research concerning the methods used by subjects to discover deception in non-laboratory settings suggests that only a more moderate form of pessimism is in order.[2]

[2] Two remarks about the strategy of my critique of Fricker: First: Because the critique turns on facts about the reliability of monitoring and blind trust and about their capacities to produce sensitive beliefs, there is a question about whether and how it makes contact with Fricker's thoroughly internalist case for local reductionism. In what follows, I take it for granted that even the internalist will admit that considerations of reliability, sensitivity, and the like are decisive in the end. I take my cue here from Goldberg and Henderson, who argue that Fricker should not be read as maintaining that "being gullible is bad *independent of* whether it conduces to false or unreliable belief" (2006, p. 605). Note that Fricker does not take issue with this characterization in her response to Goldberg and Henderson (2006c). Second: Because my critique turns on results obtained by scientific psychology, there is a question about whether and how it makes contact with Fricker's thoroughly folk psychological case for local reductionism. Fricker supports her reliance on the folk psychology of testimony with an appeal to coherentism (1994, p. 145). But coherentism does not entitle the epistemologist of testimony to rely indefinitely on the folk psychology of testimony: in light of the epistemic superiority of scientific psychology over folk psychology, even an epistemologist ignorant of the scientific psychology of testimony at the outset of her investigation must eventually render her theory consistent with the former. Note that certain of Fricker's remarks suggest that she would accept this point (2006a, p. 601).

Though Fricker's particular local reductionism serves as my target here, the argument of the paper is of more general significance. The epistemology that she develops is (as she herself emphasizes) well-supported by our "common sense" about testimony: it might even be viewed as a refinement of that common sense; and it is thus of broad appeal. A number of other authors (e.g., Henderson 2008, who cites, in addition to Fricker, Adler 1994, Faulkner 2000, and Lyons 1997) have in fact proposed similar views about the relative merits of monitoring and blind trust. Fricker's particular epistemology of testimony thus serves as a representative of a family of related, intuitively plausible epistemologies; the argument of the paper has implications for this family as a whole. I focus on Fricker's epistemology here for two reasons: first, the theory is well-developed and influential, and it thus merits a sustained critique; second, it is necessary, in order to begin to draw out the implications of the psychology of deception detection for the epistemology of testimony, to have a theoretical framework definite enough to allow the psychology to make firm contact with the epistemology, and (as will emerge below) the framework provided by Fricker's theory is well-suited for this role.

2 Testimony and the folk psychology of deception detection

As Fricker conceives of it, the reductionist/antireductionist debate concerns the source of a subject's entitlement to the beliefs which serve as her reasons for her testimonial beliefs: according to the reductionist, the beliefs in question are ordinary empirical beliefs and thus are justified in the way in which any empirical belief is justified; according to the antireductionist (e.g., Coady 1992), on the other hand, (some of) the beliefs have a special status—the subject is entitled to assume them without evidence.[3] Fricker develops her local reductionism in opposition both to antireductionism and to global reductionism. For the global reductionist (e.g., Hume 1975), the justification of testimonial beliefs proceeds wholesale: a subject's testimonial beliefs are justified because she is in a position to perform an inductive inference to the conclusion that testimony is generally reliable (that an arbitrary testimonial utterance is probably true). For the local reductionist, in contrast, the justification of testimonial beliefs proceeds piecemeal: the subject need only have positive reason for believing that a given testimonial utterance is reliable in order for a belief that she forms on its basis to be justified.[4]

[3] There is a third option (associated with Locke 1975), "pessimistic" reductionism: the pessimistic reductionist argues that if a testimonial belief is justified, then the beliefs which support it must be justified empirically (i.e., that a reduction of testimonial justification is necessary) but also that the beliefs in question cannot be justified empirically (that no reduction is possible), concluding that therefore there are no justified testimonial beliefs. I return to pessimistic reductionism in Sect. 4.

[4] The local reductionist need not (and should not attempt to) show that a subject will on every occasion of testimony have such reason: she can (and should) acknowledge that subjects' testimonial beliefs are sometimes unjustified in virtue of their lacking appropriate reasons for those beliefs. But if we take it as given that the majority of the testimonial beliefs of a normal subject are justified, then the local reductionist does need to show that on most occasions on which testimony is accepted the subject will have appropriate reasons for the resulting testimonial belief—otherwise, her view will classify too many testimonial beliefs as unjustified. Fricker, I take it, recognizes that the task for local reductionism is to show that subjects

Fricker argues that if the broad conception of testimony (on which testimony is equivalent to "serious assertions aimed at communication" (1994, p. 137) or "tellings in general" (1995, p. 396)) is correct, then we can straightaway rule out global reductionism. If the broad conception is correct, then the subject cannot possibly perform an induction to the conclusion that testimony is generally reliable, for the straightforward reason that there is nothing meaningful to be said about the general reliability of testimony: "looking for generalisations about the reliability or otherwise of testimony …, as a homogeneous whole, will not be an enlightening project. …[W]hen it comes to the probability of accuracy of speakers' assertions …testimony is not a unitary category" (1994, p. 139). If this is right, then global reductionism is a non-starter.[5]

With Fricker, I endorse the broad conception of testimony: if our concern in the epistemology of testimony is with testimonial belief-producing processes that human subjects might actually implement, with possible policies governing their doxastic responses to testimony, then the relevant category is that of tellings in general;[6] for their policies for doxastic response to testimony will govern their responses to tellings in general. But note that Fricker's point about general reliability cannot be that there is nothing meaningful to be said about the statistical frequency of accurate testimony, for her argument for the necessity of a reduction itself turns on the claim that false testimony is sufficiently frequent to make an epistemic difference. I take it that the point is, instead, that truth is not a projectible property of testimony: the idea is, roughly, that an arbitrary testimonial utterance is not probably true in the sense that a testimonial utterance, qua testimonial utterance, does not have a propensity to be true. If truth is not a projectible property of testimony, then an inference from observations of the accuracy of testimony to the conclusion that an arbitrary testimonial utterance is probably true is illegitimate; the illegitimacy of the induction is what defeats global reductionism. But it might nevertheless be the case, e.g., that testimony is probably true in the sense that as a matter of fact testimonial utterances are more frequently true than false.

For (local) reductionists and antireductionists alike, a subject's testimonial belief is justified (if it is justified) in part in virtue of her belief that the "validity conditions" of the "epistemic link" of testimony are satisfied; the subject will also have a belief that the relevant speaker gave testimony with the relevant content. Fricker identifies the validity conditions of the epistemic link of testimony with the relevant speaker's trustworthiness on the relevant occasion and with respect to the relevant topic and so holds that if her testimonial belief is justified, the subject will have a justified belief that the relevant speaker was trustworthy on the relevant occasion and with respect

Footnote 4 continued
typically have appropriate reasons, i.e., that local reductions are not just possible but generally available (1994, p. 148).

[5] For a discussion of the implications of the narrow conception of testimony (Coady 1992) for global reductionism, see Michaelian (2008b).

[6] Note that the category of tellings in general is also the relevant one for the purposes of the psychology of deception detection: deception detection research is concerned with lies told about a wide range of topics and by a wide range of speakers.

to the relevant topic.[7] Trustworthiness, as Fricker conceives of it, breaks down into sincerity (which I take to be equivalent to honesty) and competence (where if a speaker is competent with respect to P, then her belief that P is true) (1994, pp. 132, 146, 147, 1995, p. 398). These beliefs, which are individually necessary for testimonial justification, are jointly sufficient for testimonial justification: if a subject justifiedly believes that a given speaker testified that P, that that speaker was then competent with respect to P, and that she was then honest with respect to P, and if her testimonial belief that P is appropriately related to those beliefs, then the testimonial belief is justified, for "[w]hen [a speaker is both sincere and competent], her testimony is necessarily veridical. That is: that she asserted that P, and she is sincere, and competent with respect to P, logically necessitates P" (Fricker 2004, p. 117). In order to be appropriately related to her reasons, the subject's testimonial belief need not be consciously inferred from them, but it must at least be appropriately responsive to them: if the subject believes that a speaker was incompetent, e.g., then she will not form the corresponding testimonial belief (Fricker 1994, p. 157).[8]

The disagreement between the reductionist and the antireductionist concerns the source of the subject's entitlement to her beliefs about the speaker's competence and honesty. Fricker thus has little to say about the source of the subject's entitlement to her belief about what was said, taking it as given that the subject's knowledge that she has received testimony with a certain content is unproblematic; I grant this assumption.[9] The antireductionist maintains (and the reductionist denies) that the subject is entitled to assume that the speaker was trustworthy with respect to P; the antireductionist maintains (and the reductionist denies), in other words, that the subject's testimonial belief can be justified even if she trusts the speaker blindly. This entitlement or right is not absolute but rather presumptive: according to the antireductionist, "[o]n any occasion of testimony, the hearer has the epistemic right to assume, without evidence, that the speaker is trustworthy …unless there are special circumstances that defeat this presumption" (Fricker 1994, p. 125).

Fricker assumes that the relevant defeaters will be internal (doxastic), and this seems right: given that the presumptive right thesis is meant to describe a belief-producing process that subjects can actually implement, external (normative) defeaters are irrelevant here (2006c). But there remains a question about what range of defeaters can defeat the right to assume trustworthiness; depending on how we answer this question, we end up with a weaker or a stronger presumptive right thesis, one that posits a right

[7] As Gelfert emphasizes, the internalist character of Fricker's theory is evident not only in her requirement that the subject's testimonial belief be based on her belief that the speaker was trustworthy but also in her requirement that the latter belief be accessible to the subject (2009b, p. 176).

[8] But it will often be useful for heuristic purposes to think of both the blindly trusting subject and the monitoring subject as consciously inferring their testimonial beliefs from their reasons for those beliefs (to think of subjects as inferring 'P' from 'S said that P' and 'S was then trustworthy with respect to P').

[9] The strategy of my critique of her case for local reductionism does not require me to challenge the assumption; but I want nevertheless to note that it is potentially problematic. Fricker admits that her approach is legitimate only if "it is not intrinsic to the state of understanding an utterance that it compels the hearer towards belief in what she grasps as being asserted" (1994, p. 157). But there is reason (though not conclusive reason) to think that understanding a testimonial utterance does in fact tend to result automatically in the formation of the corresponding belief (Gilbert et al. 1990, 1993; Gilbert 1991, 1993).

that can be cancelled more or less easily. I will be interested here in (what I take to be) the strongest plausible presumptive right thesis, the thesis which posits a right to assume trustworthiness whenever one does not possess evidence that the speaker is untrustworthy; the right posited by this thesis is a right to assume that a speaker is trustworthy unless one antecedently possesses a doxastic defeater for the proposition that she is trustworthy.[10] This presumptive right thesis licenses a certain belief-producing process; I will say that a subject whose testimonial beliefs are produced by this process is "blindly trusting". A blindly trusting subject receives testimony that P and forms the belief that the relevant speaker testified that P; if she does not already possess defeaters for the propositions that the speaker was honest with respect to P when she testified and that she was then competent with respect to P, then (without engaging in any additional cognitive activity) she forms the corresponding beliefs and therefore forms the belief that P; otherwise, she does not form the belief that P. The presumptive right thesis thus implies that the subject enjoys a dispensation from "the requirement to monitor and assess a speaker for trustworthiness" (Fricker 1995, p. 404).

As noted at the outset, Fricker's argument against antireductionism (her argument for the necessity of a reduction) turns on the claim that the presumptive right thesis is "an epistemic charter for the gullible and undiscriminating" (1994, p. 126): "The thesis that I advocate …is that the hearer should always engage in some assessment of the speaker for trustworthiness. To believe without doing so is to believe blindly, uncritically. This is gullibility" (1994, p. 145). But there is a question about what, exactly, Fricker might mean by 'gullibility' here. Goldberg and Henderson discuss this question at some length, distinguishing among various possible senses of the term (2006, pp. 602–603). In her response to them, Fricker specifies that she has in mind approximately the ordinary language sense of the term: a subject is gullible, in this sense, "if she has a disposition or policy for doxastic response to testimony which fails to screen out false testimony" (2006c, p. 620). As Fricker notes, this is equivalent to the second sense of 'gullible' distinguished by Goldberg and Henderson: to say that a subject is gullible, in this sense, is to say that "in circumstances C, [she] is disposed to acquire a good deal of unreliable (unsafe; insensitive; etc.) testimony-based belief" (Goldberg and Henderson 2006, p. 602). Goldberg and Henderson's formulation has a pair of virtues. First, it makes explicit that gullibility is relative to an environment: a belief-producing process use of which renders a subject gullible in one environment might not render her gullible in another environment; hence in order to show that her use of a given belief-producing process renders a subject gullible, we need not only to identify certain features of the process which will entail gullibility in a certain range of environments but also to show that the environment in which she uses the

[10] This strongest plausible presumptive right thesis is weaker than the strongest possible presumptive right thesis, the thesis which posits a right "to believe that P, just on the ground that it has been asserted, whenever one does not already possess evidence showing 'P' to be false" (Fricker 1994, pp. 142–143); the right posited by this stronger thesis is a right to assume that a speaker is trustworthy unless one possesses a doxastic defeater for the proposition that is the content of her testimony. The strongest thesis is too strong to be plausible: a subject who forms her testimonial beliefs in the manner licensed by the thesis might in certain cases believe that the speaker is untrustworthy and yet go on to accept her testimony; but this does not seem like a policy that normal subjects could actually implement.

process is included in that range. Second, it makes clear that a subject might "fail to screen out false testimony" either in the sense that her testimonial beliefs have some epistemically bad modal property (whether or not they are reliably produced, i.e., produced by a process which tends to produce more true than false beliefs) or in the sense that they are unreliably produced (whether or not they have epistemically bad modal properties).

The relevance of the former point is made clear below. The latter point means that there are two possible interpretations of Fricker's argument for the necessity of a reduction: on one interpretation, it concludes that the testimonial beliefs of blindly trusting subjects are unreliably produced (and hence that they are not justified); on the other interpretation, it concludes that many of the testimonial beliefs of blindly trusting subjects have some epistemically bad modal property (and hence are not "knowledgeable").[11] There is (as far as I can see) nothing in Fricker which settles the question which of the two versions of the argument she intends. Similarly, there is nothing which settles the question whether she is concerned with sensitivity, safety, or some other modal property. Since her argument for the general availability of reductions appeals explicitly to subjects' sensitivity to signs of untrustworthiness, a version of that argument involving sensitivity [where a subject's belief that P is sensitive iff if it were not the case that P, then (if she were to form her belief using the method that she in fact uses) the subject would not believe that P (Nozick 1981)] is particularly plausible; I will therefore also read the modal version of the argument for the necessity of a reduction as involving sensitivity (rather than safety or some other epistemically significant modal property). Since the insensitivity of the beliefs produced by the blind trust process would provide a partial explanation of the unreliability of the process, the process reliability and sensitivity versions of the argument are naturally combined; I will therefore read Fricker as advancing a single argument for both conclusions. [But note that the conclusions of the argument that the blindly trusting subject's testimonial beliefs are insensitive and that they are unreliably produced (as well as the conclusions of the argument for the availability of reductions that the monitoring subject's testimonial beliefs are sensitive and that they are reliably produced) do need to be distinguished: as the argument of Sect. 3 makes clear, a subject's testimonial beliefs might attain reliability (and so qualify as justified) without attaining sensitivity (and so qualifying as knowledge).]

The following is a natural reconstruction of the argument for the necessity of a reduction:

1. Suppose that a subject trusts blindly (i.e., that if she receives and understands testimony, then, unless she happens antecedently to have doxastic defeaters for the proposition that the speaker is trustworthy, she assumes that the speaker is trustworthy and therefore accepts her testimony).
2. On many occasions of testimony, if a speaker were untrustworthy, the subject would not antecedently have doxastic defeaters for the proposition that she is trustworthy (assumption).

[11] I assume that reliable production is necessary for justification and that sensitivity—I will take this to be the relevant epistemically significant modal property—is necessary for knowledge.

3. Many of the blindly trusting subject's beliefs that speakers are trustworthy are insensitive (from 1 and 2).
4. Many of the blindly trusting subject's testimonial beliefs are insensitive (from 1 and 3).
5. Untrustworthy testimony is quite common—the blindly trusting subject will (like any other subject) potentially receive a good deal of untrustworthy (and hence false)[12] testimony (assumption).
6. The blindly trusting subject will potentially end up with many false testimonial beliefs—the blind trust process is unreliable (from 4 and 5).

It is plain that when a subject (whether blindly trusting or not) receives testimony, she will often lack prior knowledge of the speaker's (un)trustworthiness; thus premise 2 of the argument must be granted. The success of the argument thus turns on premise 5. If the argument succeeds, then Fricker has shown that the testimonial beliefs of a subject who assumes trustworthiness when she does not already have relevant defeating beliefs are epistemically defective, thus that subjects do not enjoy a defeasible right to assume trustworthiness, thus that if a testimonial belief is not epistemically defective, then it must be supported by an empirically grounded belief in trustworthiness. Fricker's argument for the general availability of reductions is designed to show that the beliefs in trustworthiness of normal subjects are indeed typically empirically grounded: normal subjects typically perform a certain sort of cognitive activity when they receive testimony; their performance of this activity ensures that their beliefs in trustworthiness are empirically grounded; and thus the testimonial beliefs of normal subjects typically are not epistemically defective.

Fricker suggests that normal subjects form their testimonial beliefs using a monitoring process. Like the blind trust process, the monitoring process takes as input the subject's beliefs about what the speaker said, about the competence of the speaker, and about the honesty of the speaker and produces as output either a testimonial belief or no belief at all: if the subject believes that the speaker asserted that P, that the speaker was then competent with respect to P, and that the speaker was then honest with respect to P, the process outputs a belief that P; otherwise (e.g., if the subject believes that the speaker was incompetent with respect to P), the process outputs no belief. Unlike the blind trust process, the monitoring process involves (unconscious (Fricker 1994, p. 150)) cognitive activity to determine honesty and competence. Fricker allows that honesty and competence are default settings within the monitoring process: should the subject have no evidence of untrustworthiness, she assumes trustworthiness. But she emphasizes that these are merely defaults within the process: the subject assumes trustworthiness only given that she has actively attempted to determine whether the speaker is trustworthy.

I am interested here primarily in what Fricker has to say concerning the determination specifically of honesty, and she suggests that the methods used by subjects to determine honesty in general differ from those used by them to determine competence. But she does describe one method which, she says, is sometimes used to determine trustworthiness as a whole (that is, to determine both honesty and competence at once):

[12] Like Fricker, I ignore the rare cases in which dishonesty and incompetence cancel each other out.

trustworthiness can sometimes, according to Fricker, be established "in the approved Humean fashion, induction from observed constant conjunction—we trust one person's report, because she has built up a track record of accuracy; we distrust another because she has accumulated the opposite" (1994, p. 135). This "track-record" method, if indeed it can be used to establish trustworthiness, can (as Fricker acknowledges) be so used only rarely (viz., when the subject has on a sufficient number of previous occasions personally verified the accuracy of the testimony of the relevant speaker on the relevant topic), and so it is unlikely that the success of Fricker's argument for the general availability of reductions will be affected by the correctness of her claim regarding it: the reliability of the monitoring process and the frequency with which it produces sensitive beliefs will not be affected significantly by the occasional use of the track-record method, and I therefore disregard it in what follows.[13]

How, then, do subjects usually determine honesty and competence? Since incompetence is typically not indicated by perceptible signs, a subject's determination of a speaker's competence will typically flow from her prior beliefs concerning the speaker's cognitive capacities (Fricker 1994, p. 150, 1995, pp. 404–405). But dishonesty is often indicated by perceptible signs (cues), and the subject's determination of a speaker's honesty will typically be a result of her monitoring the speaker for such signs (Fricker 1995, pp. 404–405): "[e]xpert dissimulators among us being few, the insincerity of an utterance is very frequently betrayed in the speaker's manner, and so is susceptible of detection by …a quasi-perceptual capacity" (Fricker 1994, p. 150). I grant the claim that subjects typically use their prior knowledge of speakers to determine their competence; I grant, moreover—the reason for which Fricker needs this will be evident—that judgements of competence are typically sensitive. My focus here is not on the monitoring process as a whole but rather on the process of monitoring for cues to dishonesty the results of which feed into that larger process.[14,15]

[13] I want nevertheless to note that there is reason to doubt that the track-record method can be used to establish trustworthiness. The suggestion that it can be so used presupposes that trustworthiness is a projectible property of the testimony of a particular speaker (on a particular topic). But in presupposing this, we run the risk of underestimating the role played by the environment in shaping the trustworthiness of testimony, that is, of committing (something like) the fundamental attribution error. Indeed, O'Sullivan argues that subjects' attempts to determine honesty, in particular, are often thwarted by their commission of the fundamental attribution error (2003). In general, note that though it is natural to suppose that past interactions with a speaker give a subject an advantage in determining her honesty, in fact it appears that this is not the case (Anderson et al. 1999a; Feeley and Young 1998).

[14] Note that both the monitoring subject and the blindly trusting subject rely on their prior knowledge to establish competence. Since the monitoring subject assumes competence if she does not have reason to think the speaker incompetent, her beliefs that speakers are competent are produced by the same process that the blindly trusting subject uses to produce her beliefs that speakers are competent: each assumes competence unless she antecedently has defeaters for the proposition that the speaker is competent. Thus it seems that if the judgements that speakers are competent of monitoring subjects are typically sensitive, so are the judgements that speakers are competent of blindly trusting subjects; the epistemic differences between the monitoring and blind trust processes (if there are any) thus must stem from the differences between them with respect to the formation of judgements that speakers are honest.

[15] Though my focus here is on monitoring for dishonesty, there are questions about the adequacy of Fricker's assumptions about monitoring for incompetence parallel to those I raise about her assumptions about monitoring for dishonesty. These questions are equally pressing but would themselves require a paper-length treatment. (See Gelfert 2009a for a brief discussion of related questions.)

Unfortunately, Fricker does not say anything systematic about the cues which, she supposes, typically betray dishonesty. I will assume that she has in mind the cues posited by the folk psychology of deception detection: this is consistent with her claim that subjects' folk psychological knowledge often enables them to detect untrustworthiness; and her explicit remarks also suggest that she has these folk psychological cues in mind.[16] As we will see, the assumption that there are cues to deception is correct; the interesting question is whether these cues are the ones posited by folk psychology.

Monitoring, as Fricker conceives of it, is a matter of having a certain counterfactual sensitivity: if a subject monitors for untrustworthiness, then "it is true throughout of [her] that if there were any signs of untrustworthiness, then she would pick them up" (1994, p. 154); thus if a subject monitors for cues to dishonesty, in particular, it is true throughout of her that if there were any cues to dishonesty, then she would pick them up. But this use of 'monitoring' is potentially misleading, for it does not permit us to speak of (what it is natural to describe as) "ineffective monitoring", an activity which attempts to detect cues to dishonesty but which (for one reason or another) fails to detect such cues when they are present. I therefore depart from Fricker's use of the term, using 'monitoring' to refer to any activity which attempts to detect cues to dishonesty, whether or not it succeeds in detecting them, and 'effective monitoring' to refer to an activity which attempts to detect cues to dishonesty and which succeeds in detecting them. As we will see, the assumption that subjects monitor for cues is correct; the interesting question is whether this monitoring is effective.

But these are not the only interesting questions: even if a subject monitors effectively, her judgements that speakers are honest will frequently be insensitive unless several further conditions are met. First, if the subject's sensitivity to cues is to translate into sensitivity to dishonesty itself, Fricker needs it that there are bona fide cues to dishonesty and that the cues to which subjects are sensitive correspond closely to these bona fide cues. Second, if the subject's sensitivity to cues is to translate into sensitivity to dishonesty itself, Fricker needs it that dishonesty is typically accompanied by cues. Finally, since what matters ultimately is that the subject's judgements that speakers are honest are sensitive (rather than that the subject herself is sensitive to dishonesty), Fricker needs it that the subject's determinations of honesty are appropriately responsive to the results of her monitoring for honesty.

Taking these requirements into account, I propose the following as a reconstruction of the argument for the general availability of reductions:

1. Suppose that a subject monitors (i.e., that she attempts to determine competence by relying on her prior knowledge of the speaker's cognitive capacities and to determine honesty by monitoring the speaker for cues to dishonesty, and that she accepts a speaker's testimony iff she believes that the speaker gave the testimony and that she was then both honest and competent with respect to the topic of her testimony).

[16] She says, e.g., that the sincerity of an utterance can be determined "through attention to such features as tone of voice, and manner of the speaker" (1994, p. 147). And she mentions, e.g., "a hesitancy in the voice" and "an insincere-seeming smile" (2004, p. 117). These cues are among the components of an apparently cross-cultural stereotype of the behaviour of liars (Global Deception Research Team 2006).

2. The subject's beliefs that a given speaker gave testimony with a certain content are frequently sensitive (assumption).
3. The subject's beliefs that a given testimonial utterance is competent are frequently sensitive (assumption).
4. The subject's monitoring for cues to dishonesty is effective relative to the cues for which she monitors (assumption).
5. There are bona fide cues to dishonesty (assumption).
6. The cues for which the subject monitors correspond closely to the bona fide cues to dishonesty (assumption).
7. The subject's monitoring for cues to dishonesty is effective relative to the bona fide cues to dishonesty (from 4, 5, and 6).
8. Dishonesty is frequently accompanied by the bona fide cues (assumption).
9. The subject is frequently sensitive to dishonesty (from 7 and 8).
10. The subject's judgements that speakers are honest are appropriately responsive to the results of her monitoring for cues to dishonesty (assumption).
11. The subject's judgements that speakers are honest are frequently sensitive (from 9 and 10).
12. If a subject has sensitive beliefs that a speaker testified that P on a certain occasion, that the speaker was then competent with respect to P, and that the speaker was then honest with respect to P, and if her belief that P (formed on the basis of the speaker's testimony) is appropriately responsive to those beliefs, then her belief that P is sensitive (from the definition of sensitivity and the fact that if a speaker is honest and competent when she testifies that P, then P).
13. The monitoring subject's testimonial beliefs are frequently sensitive (from 1, 2, 3, 11, and 12).
14. Whatever the frequency of false testimony, the monitoring subject's testimonial beliefs are frequently true—the monitoring process is reliable (from 1, 13, and the fact that if most of the beliefs produced by a process are sensitive, then most of the beliefs produced by the process are true).

Granted the additional assumption that the testimonial beliefs of normal subjects are produced by the monitoring process, the argument shows (if it succeeds) what Fricker needs to show: if a subject uses the monitoring process, then her testimonial beliefs will typically be sensitive and reliably produced; if the testimonial beliefs of normal subjects are produced by the monitoring process, this accounts for the fact that (though we do not have a right to assume trustworthiness) most of the testimonial beliefs of normal subjects are justified and knowledgeable.[17]

3 Testimony and the scientific psychology of deception detection

The suggestion that testimony and memory are broadly analogous epistemic sources is a familiar one: the basic thought is that both are purely preservative sources, in the sense

[17] This overstates the import of the argument, since it is unclear whether a belief must have modal properties in addition to sensitivity in order to qualify as knowledge.

that neither can produce new justification or knowledge.[18] Goldberg and Henderson propose a novel and more specific analogy between memory and testimony: the two sources are, they propose, analogous in terms of the role played by monitoring in each. Goldberg and Henderson suggest that the process of remembering involves a monitoring process which checks retrieved memories for various properties, including coherence with the subject's beliefs, vividness, etc. and sorts them on this basis. If the monitoring process detects that a retrieved memory is incoherent with the subject's other beliefs, e.g., then the subject does not form the corresponding memory belief. If none of the properties for which the subject monitors are detected, then the subject forms the corresponding memory belief (2006, p. 613). The monitoring process involved in remembering thus attempts to sort "trustworthy" (genuine) from "untrustworthy" (merely apparent) memories, thereby ensuring the reliability of remembering. Analogously, the process of forming a testimonial belief involves a monitoring process which checks received testimonial utterances for signs of untrustworthiness and sorts them on this basis. If the monitoring process detects signs of untrustworthiness, then the subject does not form the corresponding testimonial belief. If no signs of untrustworthiness are detected, then the subject forms the corresponding testimonial belief (2006, p. 615). The monitoring process involved in testimony thus attempts to sort trustworthy from untrustworthy testimony, thereby ensuring the reliability of accepting testimony.

The analogy is instructive not because it works but rather because of the way in which it fails. Goldberg and Henderson apparently do not base their description of the role played by monitoring in memory on the psychology of metamemory. But a review of the metamemory literature shows that their description is approximately correct: a monitoring process is in fact involved in the process of remembering; and this process is, moreover, a means of ensuring the reliability of remembering (Johnson and Raye 1981, 2000; Johnson 1988; Johnson et al. 1993; Mitchell and Johnson 2000; Kelley and Jacoby 1998; Smith et al. 2003; Wilson and Brekke 1994). What the review of the deception detection literature provided below shows, in contrast, is that while the formation of testimonial beliefs is indeed like the formation of memory beliefs in that both include monitoring processes, the similarity ends there. Remembering is reliable in part because it involves a monitoring process which is sensitive to the origins of retrieved memories in various sources (Michaelian 2008a). But the monitoring process involved in accepting testimony is not sensitive to the trustworthiness of speakers, and thus it cannot ensure reliable formation of testimonial beliefs.[19]

An initial survey of the deception detection literature bodes ill for Fricker's case for the general availability of local reductions, for it suggests that the determinations of honesty (judgements that speakers are honest and judgements that they are dishonest) of normal subjects are not reliably produced. Levine et al., e.g., remark that "the

[18] For recent challenges to the view of memory and testimony as preservative, see Lackey (1999, 2005) and Michaelian (2008a).

[19] Douven and Cuypers argue that Fricker overestimates subjects' ability to acquire evidence sufficient to ground beliefs about speakers' trustworthiness (2009). Their focus is on the (in)coherence of received testimony with the subject's background beliefs rather than on signs of (dis)honesty, but their strategy is consistent with that developed here.

belief that deception detection accuracy rates are only slightly better than fifty-fifty is among the most well documented and commonly held conclusions in deception research" (1999, p. 126). In an early review of deception detection studies, Kraut found an average accuracy rate of 57% (1980); this differs little from the more recent findings of reviews by Vrij (a 56.6% average accuracy rate) (2000) and Bond and DePaulo (a 54% accuracy rate) (2006).[20] I will assume (generously) that the average deception detection accuracy rate is about 57%.[21]

If we assume that (as seems plausible) most determinations of honesty are produced by a single process and that (as also seems plausible) the process in question produces only determinations of honesty (and not also beliefs about other topics), the 57% statistic gives us a measure of the reliability of the process. According to Fricker normal subjects typically monitor for dishonesty; if this assumption is right, then the 57% statistic gives us a measure of the reliability of the process of monitoring for dishonesty. If most of the beliefs produced by the process used by normal subjects to determine honesty were sensitive, then the process would be highly reliable; since the process is relatively unreliable, we can conclude that it is not the case that most of the determinations of honesty of normal subjects are sensitive. If Fricker is right that normal subjects typically monitor for dishonesty, we can infer from the poor reliability of the process that it is not the case that the determinations of honesty of monitoring subjects are typically sensitive. Thus the 57% statistic appears to imply that the argument for the general availability of local reductions fails: the argument infers from the intermediate conclusion that the determinations of honesty of monitoring subjects are typically sensitive (along with the assumptions that their beliefs concerning what was said are typically sensitive and that their determinations of competence are typically sensitive) that their testimonial beliefs are typically sensitive and hence that their testimonial beliefs are reliably produced; but whether because they do not monitor or because the monitoring process does not work as Fricker supposes it to work, the determinations of honesty of normal subjects are not typically sensitive.

The argument for the general availability of reductions itself suggests a number of possible explanations for the 57% statistic. First: That subjects do not monitor at all but instead rely on heuristics would, if the heuristics in question are poor or inappropriate, explain the poor reliability of the process by which determinations of honesty are produced (Vrij 2004, p. 163; Levine et al. 1999, p. 126). There is indeed a strong

[20] Note that these numbers are averages across subjects. While it is sometimes suggested that there are significant individual differences in deception detection ability, Bond and DePaulo (2008) suggest that these individual differences are in fact insignificant.

[21] A defender of monitoring might at this point raise concerns about the ecological validity of the experiments on which these findings are based. Deception detection researchers themselves have (naturally) had much to say about the ecological validity of the range of experiments in question, and a proper review of this discussion here would lead us too far afield. Instead, I will simply point out that the research on which I rely most heavily here addresses two of the main concerns about the ecological validity of deception detection experiments: the research on the role of the base rate of honesty in determining detection accuracy discussed below responds to the worry that the frequency of lying in the laboratory does not reflect the frequency of lying in everyday contexts; the research on everyday methods of lie detection discussed in Sect. 4 responds to the worry that the methods of lie detection available to subjects in laboratory settings do not correspond to the methods they use in everyday contexts. For additional discussion of the ecological validity of deception detection research, see Levine and Kim (2008) and McCornack (1997).

case to be made for the view that heuristics play an important role in accounting for the 57% statistic. But heuristics are not the whole story, for it is also clear that subjects do in fact monitor for cues to deception. As Park and her colleagues point out, the results of several distinct lines of research demonstrate that "if the verbal and nonverbal behavior of message source is the only information participants have to go on when making veracity judgments, and individuals are required to make truth-lie judgments, then they use [the speaker's verbal and nonverbal behavior]" (Park et al. 2002, p. 148); they cite research into self-reported use of cues (Feeley and Young 2000), research into correlations between actual behaviours and determinations of honesty (Stiff and Miller 1986), and research in which specific behaviours are induced in order to discover whether they affect determinations of honesty (Henningsen et al. 2000). It appears, then, that if heuristics are used, they are used within a process of monitoring for cues to dishonesty.

Second: That there are no reliable cues to deception would account for the 57% statistic, even if subjects monitor. But there are in fact some reliable cues to deception. According to Vrij, "liars have a tendency to speak with a *higher pitched voice* …; to include *fewer details into their accounts* …; to make *few illustrators* (hand and arm movements designed to modify and/or supplement what is being said verbally), and to make *fewer hand and finger movements* (non-functional movements of hands and fingers without moving the arms)" (2004, pp. 161–162). There is not perfect consensus among deception detection researchers with respect to the precise membership of the list of cues to deception.[22] But what matters for present purposes is not the precise membership of the list of cues to deception; what matters is only that there are cues to deception.

Thus there is no suggestion in the deception detection literature that Fricker's assumption that the process by which normal subjects produce their testimonial beliefs involves a process of monitoring for dishonesty is incorrect; nor is there any suggestion that there are no cues to dishonesty. But as noted above, even if subjects monitor, and even if there are cues to dishonesty, there is no guarantee that subjects' determinations of honesty will be sensitive.

Third: That the task of detecting cues is difficult would explain subjects' insensitivity to dishonesty. Vrij catalogues a number of sources of the difficulty of the task of detecting deception. For example: though there are cues to deception, these are difficult to detect because the differences between the behaviour of honest speakers and that of dishonest speakers are small (2004, p. 163); cues to deception vary from speaker to speaker, so that if a subject always relies on the same set of cues, she will often make incorrect determinations of honesty (2004, p. 166); the meaning of the cues provided by a single speaker varies from context to context, and failure to take this into account will also result in many incorrect determinations (2004, p. 167); cues to deception are not cues directly to deception itself but rather to phenomena that often accompany deception—the same phenomena sometimes accompany dishonesty, and failure to take this into account will also result in many incorrect determinations (2004, p. 170).

[22] See DePaulo et al. (2003) for a detailed review of work on cues to deception.

Fourth: That subjects monitor but monitor for the wrong cues or for too few of the right cues would also explain the 57% statistic. And though there are some bona fide cues to deception, there is little overlap between this set of cues and the set of cues for which subjects monitor (Vrij 2004, p. 164; Zuckerman et al. 1981). For example: nervousness (including gaze aversion) is part of the folk psychological stereotype of the liar; but "clear cues of nervous behaviour, such as gaze aversion and fidgeting, do not appear to be related to deception" (Vrij 2004, p. 162).

Fifth: That deception is rarely accompanied by cues would also help to account for the 57% statistic. There are some cues to deception, but not all of the cues to deception will accompany every instance of deception. And thus though there is some overlap between the cues for which subjects monitor and the bona fide cues, this is not sufficient to ensure that subjects are sensitive to dishonesty: as Feeley and Young point out, since subjects attend to only some of the bona fide cues, and since not all of the bona fide cues accompany every instance of deception, "[t]he unfortunate result is that the "hits" are dwarfed by the "misses" in deception detection" (1998, p. 114).[23]

In sum: It appears that Fricker's assumption that subjects monitor for cues to deception is correct, and it appears that her assumption that there are cues to deception is also correct. Since the process of monitoring for deception is relatively unreliable, we know that at least many of the beliefs that it produces are insensitive. We have seen that the process of monitoring for dishonesty will produce sensitive judgements that speakers are honest only given some additional assumptions. And there is reason to doubt each of those assumptions.

We can infer from the poor reliability of the process of monitoring for dishonesty that subjects' determinations of honesty are not typically sensitive. And since testimonial beliefs rely for their sensitivity in part on the sensitivity of determinations of honesty, it seems that we can infer from the fact that determinations of dishonesty are not typically sensitive that testimonial beliefs are not typically sensitive. But we cannot infer from the fact that testimonial beliefs are not typically sensitive that they are unreliably produced. The argument for the general availability of reductions attempts to show that the monitoring process is reliable by showing that it produces mostly sensitive testimonial beliefs; but the process might also be reliable because the beliefs that it receives as input are mostly true (even if insensitive).

But the 57% statistic appears to imply not only that testimonial beliefs are not typically sensitive but also that they are unreliably produced. Whether a subject monitors or trusts blindly, her reasons for her testimonial belief that P are the further beliefs that the relevant speaker testified that P, that she was then competent with respect to P, and

[23] A sixth explanation: It might be that if a subject is sensitive to dishonesty, her determinations of honesty are nevertheless insensitive. This will occur if those determinations are not appropriately responsive to the results of her monitoring: a subject might in some sense detect cues to deception but go on despite this to form the belief that the speaker is honest. Anderson and his colleagues report that there is "some very intriguing evidence that reports of cue usage vary systematically in a very important way: they discriminate truths from lies"—this, despite the fact that in their study "perceivers' explicit attempts at detecting deception …were no more accurate than chance" (Anderson et al. 1999b, p. 84). Their suggestion is that a subject's attention tends to be drawn to certain cues when the speaker is lying and to certain other cues when the speaker is honest but that her determinations of honesty are not affected by her attending to some cues rather than others (Anderson et al. 1999b, pp. 84–85). Thus there might be a sense in which though their determinations of honesty are insensitive, subjects themselves are implicitly sensitive to dishonesty.

that she was then honest with respect to P; we can thus for heuristic purposes think of both monitoring and blind trust as belief-dependent processes (processes which take beliefs as input and produce further beliefs as output) and thus as processes which are at most conditionally reliable (processes which tend to produce mostly true beliefs when given true beliefs as input) (Goldman 1979). Thus it appears that if the process which produces determinations of dishonesty is unreliable, so will be the larger process by which testimonial beliefs are produced: even if the processes responsible for the production of the other beliefs which feed into the larger process are perfectly reliable, so that they always generate true beliefs as input to the process, since determinations of honesty also feed into the larger process, the latter will be only 57% reliable (on average). In particular: if Fricker is right that determinations of honesty are normally produced by the monitoring process—and there is no evidence from the psychology of deception detection that she is wrong about this—the process will be only 57% reliable (on average). Though it is not clear what threshold of reliability a process needs to reach before it produces justified beliefs, presumably 57% reliability is insufficient for justification. [And if, e.g., the process which produces determinations of competence is only moderately reliable, the larger process (the monitoring process, if Fricker is right) might easily turn out to be much less than 57% reliable.] If this is right, then the second main conclusion of the argument for the general availability of reductions, too, is false: testimonial beliefs are not reliably produced.

But this quick inference from the 57% statistic to the conclusion that testimonial beliefs are insensitive and unreliably produced is too quick: the reasoning of the argument is confused. The 57% statistic does not by itself imply that testimonial beliefs are insensitive. Nor does it imply that testimonial beliefs are unreliably produced. Once we have identified the confusion involved in the argument, we will see that the former conclusion is true but that there is reason to hope that the latter conclusion is false.

Recall that Fricker suggests that the monitoring process includes a default setting in favour of honesty: if a subject fails to detect signs of dishonesty, then she assumes that the speaker is honest. I noted above that there is a role for the use of heuristics in accounting for the unreliability of the process which produces determinations of honesty; and there is reason to think that a heuristic similar to that suggested by Fricker plays a key role here. Because subjects are insensitive to dishonesty, the frequency with which their determinations of honesty are correct (the deception detection accuracy rate) can be predicted by a model [the Park–Levine model (Park and Levine 2001)] which makes no reference at all to monitoring and instead simply assumes that subjects prefer to judge that speakers are honest, that they are "truth-biased". That subjects are truth-biased is a robust finding: "[n]umerous studies have found that independent of actual message veracity, individuals are much more likely to ascribe truth to other's message than deceit" (Levine et al. 1999, p. 126). The Park–Levine model predicts accuracy rates on the basis simply of this truth-bias and the base-rate of honesty (the frequency of honest utterances)—the model makes no reference to monitoring for cues to deception. That subjects (in the manner suggested by Fricker) assume honesty when they fail to detect cues to dishonesty is a potential partial explanation of the truth-bias: if a subject detects "cues" relatively infrequently, and if she judges that the speaker is honest whenever she does not detect "cues", then she will frequently judge that the

speaker is honest; if she is insensitive to dishonesty, then she will tend to judge that the speaker is honest even when she is in fact dishonest.

Consideration of the Park–Levine model reveals that the 57% statistic is misleading. Initially, the model seems to imply that monitoring for dishonesty will be significantly less reliable than the statistic at first suggests. The 57% statistic is often taken to mean that subjects detect deception at a rate slightly better than chance; but the Park–Levine model calls attention to the fact that while the probability of making a correct determination of honesty (correctly judging that the speaker is honest or correctly judging that she is dishonest) is indeed slightly better than chance, the probability of correctly judging that a speaker is dishonest, in particular, is in fact worse than chance. The model is suggested by the observation that the fact that subjects are truth-biased suggests that we should expect to find a certain "veracity effect" (to find that detection accuracy is a function of message honesty): "[t]o the extent that people are truth-biased, and to the extent that chance or guessing contributes to accuracy rates, laws of probability dictate that people are more likely to correctly identify truths than lies"; in particular (if the base rate of honesty is .5, as is typical in deception detection experiments) the frequency of correct judgements that speakers are honest should be significantly greater than chance, while the frequency of correct judgements that speakers are dishonest should be significantly worse than chance (Levine et al. 1999, pp. 127–128).

Deception detection accuracy is determined by (what Park and Levine call) "truth accuracy" (the probability that the subject correctly judges that the speaker is honest) and "lie accuracy" (the probability that she correctly judges that the speaker is dishonest). Letting 'A' abbreviate 'the subject makes a correct determination of honesty', 'H' abbreviate 'the subject judges that the speaker is honest', and 'T' abbreviate 'the speaker is honest', the equation for deception detection accuracy is obviously

$$P(A) = P(H\&T) + P(\sim H\& \sim T). \tag{1}$$

The equations for truth accuracy ($P(H\&T)$) and lie accuracy ($P(\sim H\& \sim T)$), in turn, are the following[24]:

$$P(H\&T) = P(H|T) \times P(T) \tag{2}$$

$$P(\sim H\& \sim T) = P(\sim H| \sim T) \times (1 - P(T)). \tag{3}$$

The relevant conditional probabilities (the strength of the truth-bias) are determined empirically. On average, according to Levine et al. (1999), $P(H|T) = .779$ and $P(\sim H| \sim T) = .349$. Thus if the base rate of honesty is .5, the model nicely predicts the 57% statistic[25]: $P(H\&T) = .779 \times .5 = .3895$; $P(\sim H\& \sim T) = .349 \times .5 = .1745$; and thus $P(A) = .3895 + .1745 = .5640$. The model also calls attention to the fact that that statistic is less meaningful than it is normally taken to be: while the probability that a subject will make a correct determination of honesty

[24] Note that the model assumes (as I have done throughout) that a given utterance is either honest or dishonest and that the subject always judges either that the speaker is honest or that she is dishonest.

[25] Note that Levine et al. (2006) provides additional experimental confirmation of the model.

is slightly better than chance ($=.5$), the probability that she will correctly judge that a speaker is honest (her truth accuracy) is significantly higher than the probability that she will correctly judge that a speaker is dishonest (her lie accuracy); in fact, while her truth accuracy is significantly better than chance—flipping a coin would produce correct judgements that speakers are honest 1/4 of the time—her lie accuracy is significantly worse than chance.

This is at first glance bad news for Fricker; but the news is not as bad as it seems at first, for closer consideration of the model shows that lie accuracy is irrelevant to the reliability of the monitoring process. Recall that the process (like the blind trust process) takes as input the subject's beliefs that the speaker gave testimony with a certain content, that the speaker was then competent with respect to the topic of her testimony, and that she was then honest with respect to that topic and produces as output either a testimonial belief or no belief at all: if the subject believes that the speaker asserted that P, that the speaker was then competent with respect to P, and that the speaker was then honest with respect to P, the process outputs a belief that P; otherwise, the process outputs no belief—in particular, if the subject believes that the speaker was dishonest, then she does not form the corresponding testimonial belief. Hence the worse-than-chance probability of correctly judging that a speaker is dishonest does not affect the reliability of the monitoring process: when a subject judges that a speaker is dishonest, she simply does not go on to form a testimonial belief.

But neither is the better-than-chance probability of correctly judging that a speaker is honest what matters here—truth accuracy, like lie accuracy, does not affect the reliability of the monitoring process. For (obviously) it is not only correct judgements that speakers are honest which feed into the process but rather all judgements that speakers are honest. Thus what matters is not the probability of accurately judging that a speaker is honest but rather the probability that a given judgement that a speaker is honest is accurate. In other words, what matters for present purposes is neither overall deception detection accuracy ($P(H\&T) + P(\sim H\& \sim T)$) nor truth accuracy ($P(H\&T)$) but rather (what I will call) "honesty accuracy". Letting 'R' abbreviate 'the subject makes a correct judgement that the speaker is honest', the equation for honesty accuracy is[26]

$$P(R) = P(H\&T)/(P(H\&T) + P(H\& \sim T)). \tag{4}$$

Whereas the deception detection accuracy rate is the frequency with which judgements that speakers are honest and judgements that speakers are dishonest are correct, the honesty accuracy rate is the frequency with which judgements that speakers are honest, in particular, are correct. Since it is judgements that speakers are honest which feed into the larger monitoring process, it is only the latter frequency that affects the reliability of that process.

We already know the equation for $P(H\&T)$. The equation for $P(H\& \sim T)$ is the following:

[26] This is equivalent to $P(T/H)$.

$$P(H\& \sim T) = P(H| \sim T) \times (1 - P(T)) = (1 - P(\sim H| \sim T)) \times (1 - P(T)). \quad (5)$$

Thus if the base rate of honesty is .5, the model predicts an honesty accuracy rate of about .54: $P(H\&T) = .779 \times .5 = .3895$; $P(H\& \sim T) = .651 \times .5 = .3255$; and thus $P(R) = .5448$. We are thus back more or less where we started.

The 57% statistic implies that the process of monitoring for dishonesty is relatively unreliable; it thus appears to imply that it is not the case that most of the determinations of honesty of a monitoring subject are sensitive; and it thus appears to show that the argument for the general availability of reductions does not go through. The 57% statistic also appears to imply directly that the conclusion of the argument that the testimonial beliefs of monitoring subjects are reliably produced is false, for determinations of honesty feed into the larger monitoring process. This latter inference is a mistake: deception detection accuracy does not affect the reliability of the monitoring process; what matters for the reliability of the monitoring process is rather honesty accuracy, since only judgements that speakers are honest feed into the process. But honesty accuracy is even worse than deception detection accuracy. Thus the claim that the testimonial beliefs of monitoring subjects are unreliably produced seems after all to be warranted.[27] We are now also in a position to reassess the earlier argument for the claim that the 57% statistic implies that it is not the case that most of the determinations of honesty of monitoring subjects are sensitive. Since only judgements that speakers are honest feed into the monitoring process, it is only the sensitivity of judgements that speakers are honest rather than the sensitivity of determinations of honesty in general that matters for the success of the argument. And though we can infer from the 57% statistic that determinations of honesty are not typically sensitive, we cannot directly infer from that statistic that judgements that speakers are honest, in particular, are not typically sensitive. But the explanations for the 57% statistic suggest that those judgements are not typically sensitive, for they all point to the conclusion that subjects are insensitive to dishonesty. And we can infer directly from the new 54% statistic that judgements that speakers are honest are not typically sensitive, for if they were typically sensitive, then they would very frequently be true.

But the former conclusion is premature: as we see when we consider the role of the base rate of honesty (which I have neglected so far) in determining honesty accuracy, things are slightly more complicated than this. The conclusion that judgements that speakers are honest are not typically sensitive will stand, since if they were typically sensitive, then they would mostly be true, whatever the base rate. Thus I must stand by the conclusion that testimonial beliefs are typically insensitive (and hence not knowledgeable). But since the move from the insensitivity of the beliefs produced by a process to its unreliability depends on further claims about the environment (in this case: a claim about the base rate of honesty), the conclusion that testimonial beliefs are unreliably produced (and hence not justified) is premature.

[27] In fact, since the reliability of the process depends also on the frequency with which subjects' judgements that speakers are competent are correct, and since subjects will presumably sometimes mistakenly judge that speakers are competent, the model suggests that the monitoring process might even be less than 50% reliable.

The Park–Levine model predicts honesty accuracy by means of a pair of conditional probabilities (the probability that the subject judges that the speaker is honest, given that she is honest and the probability that the subject judges that the speaker is honest, given that she is dishonest) and the base rate of honesty (the probability that the speaker is honest). The discussion so far has assumed that the base rate is .5 (as it is in most deception detection experiments), and has come on this basis to a rather pessimistic conclusion about the reliability of the monitoring process. Now, Park and Levine emphasize that given the role of the base rate of honesty in determining deception detection accuracy, the standard claim that deception detection accuracy is about 57% rests on a mistake: it is meaningless to say simply that deception detection accuracy is about 57%; since the accuracy rate varies with the base rate, the claim is meaningful only if it is relativized to the base rate of .5 (Park and Levine 2001). The base rate of honesty obviously also plays a role in determining honesty accuracy (though a change in the base rate does not have the same effect on honesty accuracy as it has on deception detection accuracy): thus it is meaningless to say simply that honesty accuracy is about 54%; since the accuracy rate varies with the base rate, the claim is meaningful only if it is relativized to the base rate of .5.[28]

If the base rate of honesty of .5 that is usually used in deception detection experiments does not reflect the frequency of honesty outside the laboratory, then the claim that our deception detection accuracy is about 57% is misleading: if the base rate is significantly higher, then the deception detection accuracy rate might be significantly better; if the base rate is significantly lower, then the deception detection accuracy rate might be significantly worse. Similarly: if the base rate of honesty is significantly higher than .5, then the monitoring process might be turn out to be quite reliable; and if the base rate is significantly lower, then the monitoring process might turn out to be quite unreliable. Anderson et al. remark that "[t]he truth "bias" may simply be an accurate reflection of ordinary life. Because people really do tell the truth far more often than they lie, it is entirely appropriate for perceivers to develop an inclination to see other people's communications as truthful" (1999a, p. 395). If it is true both that subjects frequently judge that testimony is honest and that testimony frequently is honest, then deception detection accuracy might after all be quite good, for "[t]he Park–Levine model would predict very high levels of overall accuracy under conditions of elevated truth-bias and predominantly honest communication" (Levine et al. 2006, p. 255); similarly, as we will see, the Park–Levine model predicts high levels of honesty accuracy under the same conditions.[29]

[28] I noted in Sect. 2 that even if we endorse Fricker's claim that there is nothing meaningful to be said about the probability that an arbitrary testimonial utterance is trustworthy, nevertheless we should still admit that there is a fact of the matter about the frequency of trustworthy testimony. And this is all that is required if we are to ask about the effect of the base rate of honesty on the reliability of the monitoring process.

[29] One possible explanation of the origins of this harmonious arrangement is provided by Reid's suggestion that we have a disposition "to confide in the veracity of others, and to believe what they tell us" which is matched by a disposition "to speak the truth, and to use the signs of language, so as to convey our real sentiments" (Reid 1970, pp. 238–240): subjects frequently judge that testimony is honest because they have a disposition to do so; while testimony frequently is honest because speakers have a disposition to give honest testimony. (On the emergence of these dispositions, see Ekman 1996; Michaelian 2008b.) A related explanation: Levine et al. (2008b) argues that subjects obey a "principle of veracity" (Bok 1999), that is, that

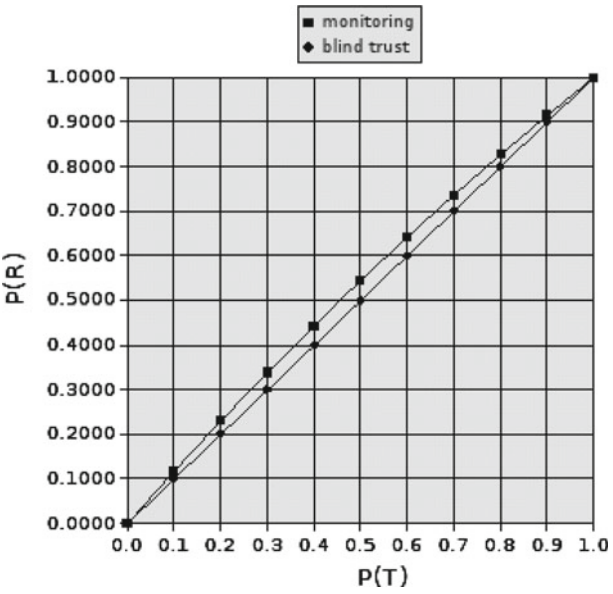

Fig. 1 Honesty accuracy of monitoring and blind trust

Figure 1 shows the honesty accuracy ($P(R)$) of monitoring subjects at different base rates of honesty ($P(T)$). The good news for the conclusion of the argument for the general availability of reductions is that if the base rate of honesty is sufficiently high, then honesty accuracy is good. Thus if the base rate of honesty is sufficiently high, then [assuming (as we have been doing all along) that the subject's beliefs about what was said and her judgements that speakers are competent with respect to the topics of their testimony are (sensitive and therefore) true sufficiently often] the monitoring process as a whole will be reliable. I return to the question of the actual base rate of honesty below.

But there is a catch: if the base rate of honesty is sufficiently high for the monitoring process to be reliable, then it is sufficiently high to render the blind trust process reliable. Because the judgements that speakers are honest of monitoring subjects are insensitive, the frequency with which those judgements are true is determined by the strength of the truth-bias of monitoring subjects in conjunction with the base rate of honesty. The argument for the necessity of a reduction shows in effect that precisely the same thing is true of blindly trusting subjects: because their judgements that speakers are honest are insensitive, the frequency with which those judgements are true is determined by the strength of the truth-bias of blindly trusting subjects in conjunction with the base rate—blind trust can, in fact, be seen in part as just an extreme truth-bias. And this means that there is much less difference between the frequency with which the judgements that speakers are honest of blindly trusting subjects are true and the

Footnote 29 continued
honesty is our default; Levine et al. (2008a) argues that subjects assume that others also obey the principle of veracity.

frequency with which the judgements that speakers are honest of monitoring subjects are true than we intuitively expect.

A remark of O'Sullivan's suggests that the overall deception detection accuracy of blind trust will be superior to that of monitoring if the base rate is sufficiently high: "Assume, for example, that people are honest 90% of the time. Given 90% honesty, the best heuristic is to assume that most people are honest most of the time. Then, one is wrong only 10% of the time" (2003, p. 1325). As the base rate of honesty increases, the advantage of our actual truth-bias increases. But as the base rate increases, the advantage of an even stronger truth-bias increases more rapidly. Thus if the base rate of honesty is, say, .9, simply assuming that speakers are honest is a significantly more reliable policy for the production of determinations of honesty than is monitoring speakers for cues to dishonesty. On the other hand, if the base rate of honesty is .1, then monitoring is significantly more reliable than blind trust.

Though blind trust is sometimes more reliable and sometimes less reliable than monitoring, blind trust and monitoring are, as shown in Fig. 1, virtually on a par with respect to honesty accuracy. It is natural to suppose that for a blindly trusting subject, $P(H|T) = 1$ and $P(\sim H| \sim T) = 0$. In fact, since the blindly trusting subject assumes honesty only when her prior knowledge does not rule it out, this is an exaggeration. But since it is likely that prior knowledge enables one to detect dishonesty only in a small number of cases,[30] I will assume for the sake of simplicity that these are the right probabilities. The moral of the story is that we are so bad at detecting deception that when it comes to deciding whether or not to accept testimony, we will do just as well by simply assuming that the speaker is honest. (To make the point vivid, note that the honesty accuracy of the blind trust policy is the same as the honesty accuracy of the policy of guessing: if one randomly guesses whether a speaker is honest, then one's judgements that speakers are honest will be right 100% of the time if speakers are always honest, 90% of the time if speakers are honest 90% of the time, etc.)

Given that monitoring subjects and blindly trusting subjects will detect incompetence about as frequently as each other (since both rely in this on their prior knowledge of speakers' cognitive capacities), the implication is that the monitoring and blind trust processes as a whole will be more or less equivalent in terms of reliability. To the extent that the epistemic merit of a process depends on its reliability, then, there is little to choose between them. And if we are concerned not only with the reliability of a process but also with its power [its ability to produce many beliefs (Goldman 1986)], the blind trust process might well turn out to be epistemically superior to the monitoring process: since the blindly trusting subject virtually always judges that speakers are honest, she will have more testimonial beliefs than the monitoring subject, who less often judges that speakers are honest. What matters centrally in the context of the reductionist/antireductionist debate is the process that subjects actually use to form their testimonial beliefs. But to the extent that we are interested in the ameliorative

[30] Note that prior knowledge enabled the subjects in the study described in Park et al. (2002) to detect deception only in 2.1% of the cases they reported.

question how we should form our testimonial beliefs, the point matters: it might be that we would do better than we do in fact by simply trusting speakers blindly.[31]

Neither monitoring for dishonesty nor assuming honesty typically produces sensitive judgements that speakers are honest; hence neither the monitoring process nor the blind trust process typically produces sensitive testimonial beliefs. Despite this, either process will reliably produce true testimonial beliefs if the base rate of honesty is sufficiently high. Thus the answer to the question whether our testimonial beliefs are justified turns on the answer to the question what is the base rate of honesty. It would thus be premature to conclude at this stage that our testimonial beliefs are justified, for relatively little is known about the base rate of honesty—there are only a few studies on this topic, and their results are mixed. In an early study, Turner et al. argued that lying is quite frequent (1975). In a more recent study, DePaulo and her colleagues suggest that lying (while an everyday occurrence) is relatively infrequent: the subjects in their study on average reported telling only one or two lies per day (1996).[32]

The best-case scenario is that the conclusion of the argument for the general availability of reductions that the testimonial beliefs of monitoring subjects are reliably produced can be saved. But if that conclusion is true, it is true not because the monitoring process typically produces sensitive testimonial beliefs but rather because the base rate of honesty is high. Hence though we may hope that our testimonial beliefs are reliably produced (and therefore justified), there seems to be little hope that they are sensitive (and therefore knowledgeable). Thus we seem to be led inevitably to a pessimistic reductionism: if subjects trust blindly, then their testimonial beliefs will be insensitive, so that a reduction of testimonial knowledge is necessary; but subjects' testimonial beliefs will be insensitive even if they monitor, so that reductions of testimonial knowledge are unavailable.

4 Towards a moderately pessimistic epistemology of testimony

Epistemologists of testimony typically implicitly assume that a subject acquires testimonial justification and knowledge (if she does acquire testimonial justification and knowledge) at the time at which her testimonial belief is formed, in other words, that a testimonial belief is justified and knowledgeable (if it is justified and knowledgeable) when it is initially produced; or at least they typically focus on the epistemic status of testimonial beliefs when they are initially produced. But the assumption is not inevitable; and the focus is not clearly the right one. Lackey has recently emphasized that the epistemic status of a belief held in memory can change over time as the subject gains or loses defeaters for the belief: a belief that was undefeated and therefore warranted might become defeated and therefore unwarranted when the subject acquires

[31] Another possibility is that we can improve the reliability of a subject's monitoring process by training her better to detect cues, so that more of her judgements that speakers are honest will be counterfactually sensitive. But the prospects for improvement via this route appear to be limited (DePaulo and Pfeiffer 1986; Ekman and O'Sullivan 1991; Feeley and Young 1998).

[32] Note that Serota and Levine (2008) argue that the average or one or two lies per day is potentially misleading: their study suggests that there are significant individual differences here, with some individuals telling lies frequently and many individuals telling lies only rarely.

a new defeater for it; a belief that was defeated and therefore unwarranted might become undefeated and therefore warranted when the subject loses her defeaters for it or acquires defeaters for those defeaters (2005).[33] I suggest that we should similarly admit that the epistemic status of a belief held in memory can change over time as the epistemically significant modal properties of the belief change: a belief that was sensitive and therefore knowledgeable might become insensitive and therefore not knowledgeable; a belief that was insensitive and therefore not knowledgeable might become sensitive and therefore knowledgeable. The point is distinct from but related to Lackey's: in cases of the sort she discusses, a subject gains or loses defeaters in the actual world; in cases of the sort I have in mind, the subject gains or loses defeaters in relevant possible worlds.

I have argued that the psychology of deception detection shows us that judgements that speakers are honest are typically insensitive when initially formed and thus that testimonial beliefs are typically insensitive when initially formed. This suggests a severe form of pessimism in the epistemology of testimony. But since the epistemically significant modal properties of a belief might change over time, the possibility remains that judgements that speakers are honest (though initially insensitive) might later on become sensitive and thus that testimonial beliefs (though initially insensitive) might later on become sensitive. And recent work in the psychology of deception detection gives us reason to hope that in many cases this is precisely what happens. This suggests only a moderate form of pessimism in the epistemology of testimony. The position that I have in mind is as follows. Subjects do indeed use the monitoring process to form their testimonial beliefs. Because the base rate of honesty is high, most of their beliefs that speakers are honest are true, and thus their testimonial beliefs are reliably produced and therefore justified. But because subjects are (despite their monitoring for deception) not sensitive to deception, their beliefs that speakers are honest are insensitive, and thus their testimonial beliefs are, when initially formed, insensitive and therefore not knowledgeable. But in many cases, a subject's judgement that a speaker is honest becomes sensitive at some later time, so that her testimonial belief then becomes sensitive. This is possible because though their monitoring for cues does not typically render subjects sensitive to deception, their receptivity to other sorts of evidence of deception does in many cases render them sensitive to deception in the long run (though it does not usually render them sensitive to deception at the time of its occurrence).

Sensitivity to cues to deception is only one possible means of ensuring that a subject's belief that a speaker is dishonest is sensitive. I noted above that in a small number of cases, the subject's prior evidence will also render her sensitive to deception, so that her belief that the speaker is honest will be sensitive. But in most cases, a subject's belief that a speaker is honest will be insensitive when it is initially formed: it will not be true of her that if the speaker were to have been dishonest, then she would not have believed that the speaker was honest. But if, after having judged that a speaker is honest, the subject later acquires evidence that the speaker was dishonest, she will abandon her earlier belief that the speaker was honest. Thus if it becomes true of the

[33] Lackey argues on this basis that memory is a generative epistemic source; see Senor (2007) and Michaelian (2008a) for responses to this strategy.

subject at some time after the formation of a judgement that a speaker is honest that if the speaker were to have been dishonest, then the subject would by then have received evidence that the speaker was dishonest, it becomes true of her at that time that if the speaker were to have been dishonest, then the subject would not believe that she was honest. If this occurs, her judgement that the speaker was honest becomes sensitive, and then so does her testimonial belief. When this occurs, a testimonial belief that initially was insensitive later on becomes sensitive. I suggest that this happens sufficiently often to allow us to avoid the severely pessimistic conclusion that there is virtually no testimonial knowledge: though it is surely not be true of every initially insensitive testimonial belief that it becomes sensitive at some later time, this is true of enough testimonial beliefs to save a significant portion of the testimonial knowledge that we ordinarily take ourselves to have.

This moderately pessimistic epistemology of testimony is inspired by some recent research conducted by Park et al. (2002). Their research suggests that (contrary to the implicit assumption of most deception detection research—and, we might add, contrary to Fricker's implicit assumption) subjects normally detect deception (when they do detect deception) not by monitoring for cues to deception but rather by using third-party information, physical evidence, etc. and that subjects normally detect deception (when they do detect deception) not at the time at which it occurs but rather somewhat later (Park et al. 2002). Use of cues to deception was reported in only slightly more than 11% of cases; cues by themselves were sufficient for discovering deception in only slightly more than 2% of cases. In contrast, third-party information was reported in slightly more than 52% of cases and physical evidence was reported in almost 31% of cases; third-party information by itself was sufficient in 32% of cases, while physical evidence by itself was sufficient in 18% of cases. Deception was detected at the time of its occurrence in less than 15% of cases; most lies were detected considerably after the fact. The study has serious limitations (as its authors readily acknowledge). But its results are suggestive: it seems that the way in which subjects discover deception is much more similar to the way in which a police detective discovers deception (after the fact, by relying on evidence from various sources) than it is to the way in which a polygraph discovers deception (at the time of its occurrence, by relying on behavioural cues).

If this is right, then we should expect that in the typical case, a subject's judgement that a speaker is honest is sensitive, if it is, not because if the speaker were to have been dishonest, then the subject would have picked up cues to dishonesty but rather because if the speaker were to have been dishonest, then the subject would have received other evidence indicating dishonesty (physical evidence, third-party information, etc.). And if this is right, then we should expect that in the typical case, a subject's judgement that a speaker is honest becomes sensitive, if it does, not at the time at which it is initially formed but rather at some later time. In certain cases, it becomes increasingly likely, as time passes, that evidence of dishonesty will eventually be received by the subject. And thus in certain cases in which a subject judges correctly that a speaker is honest, it eventually becomes true of her that if the speaker were to have been dishonest, then she would not believe that the speaker was honest, for, as time passes, the number of nearby worlds in which the speaker was dishonest but in which the subject has not yet received evidence of her dishonesty (and so continues to believe that she was honest)

decreases until a threshold is crossed. In cases of this sort, a judgement that a speaker is honest eventually becomes sensitive, though it was not initially sensitive. And thus in cases of this sort, a testimonial belief eventually becomes sensitive, though it was not initially sensitive.

Since we do not know how common it is for lies eventually to be discovered on the basis of evidence received after the time of the lie, we have no way of estimating in what fraction of cases an initially insensitive testimonial belief later becomes sensitive; thus we have no way of knowing how much of the testimonial knowledge we normally take ourselves to have is saved by the proposed view.[34] But we can at least say something about the sorts of testimonial knowledge that are and are not likely to be saved by the view. There is a spectrum of possible cases here: at one extreme, there is testimony about trivial matters, delivered by a speaker with whom the subject will never interact again and about whom the subject will never receive any additional information; at the other extreme, there is testimony about important matters, delivered by a speaker with whom the subject is likely to have further interactions and about whom she is likely eventually to receive additional information. In most cases at the former end of the spectrum, the testimonial belief never achieves sensitivity. But in many cases at the latter end of the spectrum, the testimonial belief eventually achieves sensitivity. The proposed view is thus an improvement over the unacceptably severe pessimism according to which there is no testimonial knowledge.[35]

Acknowledgements Thanks to Louise Antony, Charles Clifton, Jeff Dunn, Axel Gelfert, Namjoong Kim, Hilary Kornblith, Tim Levine, Jonathan Schaffer, Joseph Shieber, and two anonymous referees for discussion and comments.

References

Adler, J. (1994). Testimony, trust, knowing. *Journal of Philosophy, 91*, 264–275.

Anderson, D., Ansfield, M., & DePaulo, B. (1999a). Love's best habit: Deception in the context of relationships. In P. Philippot, R. Feldman, & E. Coats (Eds.), *The social context of nonverbal behavior* (pp. 372–410). Cambridge: Cambridge University Press.

Anderson, D., DePaulo, B., Ansfield, M., Tickle, J., & Green, E. (1999b). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? *Journal of Nonverbal Behavior, 23*, 67–89.

Begg, I., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*, 446–458.

Bok, S. (1999). *Lying: Moral choice in public and private life*. New York: Vintage.

Bond, C., & DePaulo, B. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*, 214–234.

Bond, C., & DePaulo, B. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*, 477–492.

[34] An additional complication: As Schacter et al. (2000) point out, semantic memory generally does not preserve information about the source of a belief. The poverty of source memory suggests that in many cases, learning about the dishonesty of a source after the fact will have no impact on beliefs formed by relying on that source's testimony; Begg et al. (1992) suggests something like this possibility. Indeed, a "sleeper effect", in which testimony from a discredited source later comes to be accepted, sometimes occurs (Kumkale and Albarracín 2004) (perhaps because the subject loses memory for the source of the testimony over time, or perhaps because the memories become dissociated).

[35] After this paper was written, I learned that Joseph Shieber has an unpublished paper in which he also begins to bring the psychology of deception detection into contact with the epistemology of testimony.

Coady, C. (1992). *Testimony: A philosophical study*. Oxford: Oxford University Press.
DePaulo, B., Kashy, D., Kirdendol, S., Wyer, M., & Epstein, J. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979–995.
DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118.
DePaulo, B., & Pfeiffer, R. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16*, 249–267.
Douven, I., & Cuypers, S. (2009). Fricker on testimonial justification. *Studies in History and Philosophy of Science, 40*, 36–44.
Ekman, P. (1996). Why don't we catch liars. *Social Research, 63*, 801–817.
Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913–920.
Faulkner, P. (2000). The social character of testimonial knowledge. *Journal of Philosophy, 97*, 581–601.
Feeley, T., & Young, M. (1998). Humans as lie detectors: some more second thoughts. *Communication Quarterly, 46*, 109–126.
Feeley, T., & Young, M. (2000). Self-reported cues about deceptive and truthful communication: The effects of cognitive capacity and communicator veracity. *Communication Quarterly, 48*, 101–119.
Fricker, E. (1987). The epistemology of testimony. *Proceedings of the Aristotelian Society, Supplementary, 61*, 57–83.
Fricker, E. (1994). Against gullibility. In B. Matilal & A. Chakrabarti (Eds.), *Knowing from words* (pp. 125–161). Dordrecht: Kluwer.
Fricker, E. (1995). Telling and trusting: reductionism and anti-reductionism in the epistemology of testimony. *Mind, 104*, 393–411.
Fricker, E. (2002). Trusting others in the sciences: a priori or empirical warrant? *Studies in History and Philosophy of Science, 33*, 373–383.
Fricker, E. (2004). Testimony: knowing through being told. In I. Niiniluoto, M. Sintonen, & J. Woleński (Eds.), *Handbook of epistemology* (pp. 109–130). Dordrecht: Kluwer.
Fricker, E. (2006a). Second-hand knowledge. *Philosophy and Phenomenological Research, 73*, 592–618.
Fricker, E. (2006b). Testimony and epistemic autonomy. In J. Lackey & E. Sosa (Eds.), *The epistemology of testimony* (pp. 225–253). Oxford: Clarendon.
Fricker, E. (2006c). Varieties of anti-reductionism about testimony—a reply to Goldberg and Henderson. *Philosophy and Phenomenological Research, 72*, 618–628.
Gelfert, A. (2009a). Learning from testimony: Cognitive cultures and the epistemic status of testimony-based beliefs. In T. Botz-Bornstein (Ed.), *Culture, nature, memes: Dynamic cognitive theories*. Newcastle: CSP (forthcoming).
Gelfert, A. (2009b). Indefensible middle ground for local reductionism about testimony. *Ratio, 22*, 170–190.
Gilbert, D. (1991). How mental systems believe. *American Psychologist, 46*, 107–119.
Gilbert, D. (1993). The assent of man: Mental representation and the control of belief. In D. Wegner & J. Pennebaker (Eds.), *Handbook of mental control* (pp. 57–87). Englewood Cliffs: Prentice-Hall.
Gilbert, D., Malone, P., & Krull, D. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*, 601–613.
Gilbert, D., Tafarodi, R., & Malone, P. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*, 221–233.
Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology, 37*, 60–74.
Goldberg, S., & Henderson, D. (2006). Monitoring and anti-reductionism in the epistemology of testimony. *Philosophy and Phenomenological Research, 72*, 600–617.
Goldman, A. (1979). What is justified belief? In G. Pappas (Ed.), *Justification and knowledge* (pp. 1–23). Dordrecht: Reidel.
Goldman, A. (1986). *Epistemology and cognition*. Cambridge: Harvard University Press.
Henderson, D. (2008). Testimonial beliefs and epistemic competence. *Noûs, 42*, 190–221.
Henningsen, D., Cruz, M., & Morr, C. (2000). Pattern violations and perceptions of deception. *Communication Reports, 13*, 1–9.
Hume, D. (1975). In P. H. Nidditch & L. A. Selby-Bigge (Eds.), *An enquiry concerning human understanding*. Oxford: Oxford University Press.
Johnson, M. (1988). Discriminating the origin of information. In T. Oltmanns & B. Maher (Eds.), *Delusional beliefs* (pp. 34–65). New York: Wiley.
Johnson, M., Hashtroudi, S., & Lindsay, D. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.
Johnson, M., & Raye, C. (1981). Reality monitoring. *Psychological Review, 88*, 67–85.

Johnson, M., & Raye, C. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 35–86). Cambridge: Harvard University Press.

Kelley, C., & Jacoby, L. (1998). Subjective reports and process dissociation: fluency, knowing, and feeling. *Acta Psychologica, 98*, 127–140.

Kraut, R. (1980). Humans as lie detectors: some second thoughts. *Journal of Communication, 30*, 209–216.

Kumkale, G., & Albarracín, D. (2004). The sleeper effect in persuasion: a meta-analytic review. *Psychological Bulletin, 130*, 143–172.

Lackey, J. (1999). Testimonial knowledge and transmission. *Philosophical Quarterly, 49*, 471–490.

Lackey, J. (2005). Memory as a generative epistemic source. *Philosophy and Phenomenological Research, 70*, 636–658.

Levine, T., & Kim, R. (2008). Some considerations for a new theory of deceptive communication. Unpublished.

Levine, T., Kim, R., & Blair, J. (2008a). (In)accuracy at detecting true and false confessions and denials: An initial test of a projected motive model of veracity judgments. Unpublished.

Levine, T., Kim, R., & Hamel, L. (2008b). People lie for a reason: Three experiments documenting the principle of veracity. Unpublished.

Levine, T., Kim, R., Park, H., & Hughes, M. (2006). Deception detection accuracy is a predictable linear function of message veracity base-rate: a formal test of Park and Levine's probability model. *Communication Monographs, 73*, 243–260.

Levine, T., Park, H., & McCornack, S. (1999). Accuracy in detecting truths and lies: documenting the "veracity effect". *Communication Monographs, 66*, 125–144.

Locke, J. (1975). *An essay concerning human understanding*. Oxford: Clarendon.

Lyons, J. (1997). Testimony, induction and folk psychology. *Australasian Journal of Philosophy, 75*, 163–178.

McCornack, S. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In J. Greene (Ed.), *Message production: Advances in communication theory* (pp. 91–126). Mahwah: LEA.

Michaelian, K. (2008a). Generative memory. Unpublished.

Michaelian, K. (2008b). Testimony as a natural kind. *Episteme, 5*, 180–202.

Mitchell, K., & Johnson, M. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. Craik (Eds.), *Oxford handbook of memory* (pp. 175–195). Oxford: Oxford University Press.

Nozick, R. (1981). *Philosophical explanations*. Oxford: Oxford University Press.

O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Personality and Social Psychology Bulletin, 29*, 1316–1327.

Park, H., & Levine, T. (2001). A probability model of accuracy in deception detection experiments. *Communication Monographs, 68*, 201–210.

Park, H., Levine, T., McCornack, S., Morrison, K., & Ferrara, M. (2002). How people really detect lies. *Communication Monographs, 69*, 144–157.

Reid, T. (1970). In T. Duggan (Ed.), *An inquiry into the human mind*. Chicago: University of Chicago Press.

Schacter, D., Wagner, A., & Buckner, R. (2000). Memory systems of 1999. In E. Tulving & F. Craik (Eds.), *Oxford handbook of memory* (pp. 627–643). Oxford: Oxford University Press.

Senor, T. (2007). Preserving preservationism: A reply to Lackey. *Philosophy and Phenomenological Research, 74*, 199–208.

Serota, K., & Levine, T. (2008). The prevalence of deception in American life. Unpublished.

Smith, J., Shields, W., & Washburn, D. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences, 26*, 317–373.

Stiff, J., & Miller, G. (1986). "Come to think of it …": Interrogative probes, deceptive communication, and deception detection. *Human Communication Research, 12*, 339–357.

Turner, R., Edgley, C., & Olmsteead, G. (1975). Information control in conversations: honesty is not always the best policy. *Kansas Journal of Sociology, 11*, 69–89.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. New York: Wiley.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*, 159–181.

Wilson, T., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117–142.

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior, 6*, 105–114.