

## Unreliable Testimony

Elizabeth Fricker

forthcoming in *Alvin Goldman and his Critics*,

Hilary Kornblith and Brian McLaughlin (eds.), Blackwell.

Reliabilism is the dominant theory in contemporary analytic epistemology. This would not be so without the work of Alvin Goldman. Pioneering articles, now classics, argued vigorously and persuasively for his versions of reliabilism about knowledge, and about epistemic justification (Goldman 1976, Goldman 1979). They were followed by a major book developing his new process-reliabilist epistemology (Goldman 1986). More recently Goldman has extended his reliabilist theory, now in the modified guise of Veritism, to the developing field of social epistemology including testimony of which he is – once again – a pioneering champion (Goldman 1999). If there were a philosophy Olympics Alvin Goldman would be a multi-gold medallist<sup>1</sup>.

Since the classic (Goldman 1976) several authors have proposed broadly reliabilist accounts of knowledge that differ on details. (Dretske 1981, Nozick 1981, Zagzebski 1996, Greco 2000, Sosa 2007) The core idea is that knowledge is belief that is non-accidentally or reliably true. Goldman (1976) proposed that perceptual belief is knowledge just if the believer would not form a similar but false belief in various ‘relevant alternative’ situations, thus proposing reliability conceived modally as necessary for knowledge<sup>2</sup>. Nozick (1981) recognised that, to avoid intuitively incorrect results, this modal reliability condition must be relativized to the method or process by which the belief subject to this condition is formed. Debate has focussed on the tricky and crucial ‘generality problem’ of how these methods or processes should be individuated; and on how to select the domain of ‘relevant alternative’ situations with respect to which the reliability of a belief, or belief-producing process, is evaluated. Some have maintained that these problems admit of no convincing general solution, so that reliabilism offers no informative reduction, nor even a genuinely constraining conceptually independent necessary condition for knowledge (Williamson 2000). The leading positive objection to reliabilism is that it fails to accommodate ‘internalist’ intuitions requiring an epistemic rationality or doxastic-responsibility condition on knowledge. The

---

<sup>1</sup> This paper was first drafted whilst the Olympic games in London were taking place in August 2012, rendering such parallels cognitively salient to its author.

<sup>2</sup> Goldman 1979 focusses on epistemic justification, arguing that a belief is justified just if it is produced by a reliable process. However, since one can have a false justified belief, reliability of processes that confers justification is explicated in terms of a high, but less than 100 percentage of truth of the resulting beliefs in a given world or worlds, rather than being explicated modally, as for knowledge.

internalist invokes, providing examples, the possibility of a belief-producing process possessed by an agent which is in fact perfectly reliable, while this fact remains wholly unrepresented within the agent's own epistemic perspective. The indicated conclusion is that reliability in the process by which a belief is formed is not sufficient for that belief's epistemic rationality, nor therefore for its being justified or constituting knowledge (BonJour 1980, BonJour 1985)

In contrast with this controversy over whether reliability – aptly explicated – is sufficient for knowledge, it is now widely accepted that reliability is a necessary condition for knowledge, even if not an independently well-determined one. This thesis is now hardly more controversial than the tenet that knowledge that P requires belief that P.

But consideration of testimony as a source of knowledge gives cause for some further thought on this matter. So shall I argue in the present contribution. In (Fricker 1987) I argued that the case of testimony reveals the insufficiency of reliabilism as an account of knowledge, because it cannot account for the datum that language-use is a single skill: the cognitive package that is needed and suffices for the acquisition of knowledge through testimony is also sufficient for giving testimony. This is so, I argued, because the roles of testifier and recipient respectively both require mastery of the basics of folk linguistics – a conception of the nature of the speech act of telling, and its consequent epistemic force as a source of knowledge. In the present essay, in contrast, I raise and review some considerations which throw doubt on the widely-accepted thesis (call it **R-NEC**) that reliability is necessary for knowledge.

What follows has four main parts. In sect.1 I formulate our problem of interest precisely. In sect. 2 I consider whether the generally pessimistic results in the experimental literature from social psychology concerning subjects' ability in a test situation to tell, from behavioural cues, whether a speaker is lying, present a severe challenge for R-NEC. My conclusion is that these results, interesting though they are, should not precipitate epistemologists into a panic retreat from R-NEC on the basis that it is inconsistent with what we know about situated human capacities. The experimental studies do not provide any strong evidence against the positive empirical claim that testimony recipients typically utilise background social knowledge to screen effectively for competence and honesty in the testifiers they actually encounter in everyday situations. In sect.3 I argue further that even if the positive empirical claim were shown false, this need not indicate rejecting R-NEC, as opposed to maintaining that givers and receivers of testimony jointly need to do better. In sect. 4 I develop a more classic line of thought invoking intuitions about cases, which suggests that there is however cause from another source to put in question R-NEC, at least in the case of testimonial knowledge. My conclusion is not that the core reliabilist intuition is faulty and should be abandoned, but that it needs to be accommodated in an alternative version of

reliabilism. I introduce and defend this variant theory, **Approved-List Reliabilism**. The proposal is similar to one made in (Goldman 1992) in relation to epistemic justification; here I propose it as systemising ordinary language intuitions about knowledge. In my final section 5, I draw out some threads from what has emerged so far, to show how testimony reveals the need for a new kind of analysis, illustrating the need and scope for social epistemology.

## **1. Testimony and Reliability**

**R-NEC** is the thesis that reliability (aptly explicated) of a subject's belief is necessary for that belief to be knowledge. R-NEC is grounded in a strong anti-luck intuition: if a belief amounts to knowledge, it cannot be just a matter of luck, or chance, that that belief turned out true. Goldman (1976) takes up the anti-luck intuition about knowledge and explicates it modally, in his account of perceptual knowledge: a perceptually-based belief amounts to knowledge just if there is no relevant alternative possible situation in which the subject would have formed an equivalent percept and similar belief which was false. I think this modal conception of reliability best explicates our anti-luck intuition about knowledge<sup>3</sup>. A plausible modal reliability condition on knowledge must apply the modal reliability requirement to the method or process by which the belief in question is formed<sup>4</sup>. So we can explicate reliability, as featuring in R-NEC, in the following definition:

**Reliability** (first formulation): A person who forms<sup>5</sup> a certain belief B by means of a method M she possesses, satisfies **Reliability** with respect to B just if: B is true, and there are no relevant alternative possible situations in which S forms a relevantly-similar false belief by using M.

Which alternative possible situations are the relevant ones, as invoked in **Reliability**? The objection previously mentioned might be invoked: that we have no grip on what these are, in some example, independent of our grip on whether the belief at issue amounts to knowledge or not. But we can explicate the notion of 'relevant alternative' thus: the relevant alternatives are those that *might easily* have obtained. Here, I suggest, we have an ordinary language notion that there is some independent purchase on. (You arrive in your seat at the Royal Opera House just as the doors are closing and the house-lights going down, having allowed

<sup>3</sup> We do, I think, indeed find ourselves with a strong intuition favouring the modal anti-luck requirement for knowledge when presented with some examples; Goldman [1976] tapped into this. However, my argument in sect.4 proposes that this intuition is not stable: in other examples we reject it as ruling out what are, intuitively, cases of knowledge.

<sup>4</sup> This point is implicit in Goldman (1976), and is explicit in the classic development of a reliabilist account of knowledge in Nozick (1981).

<sup>5</sup> Strictly: forms, or sustains, still on basis of her original source. There are differences crucial for some issues between epistemic dynamics versus statics, but these issues are not explored here.

only 40 minutes for your journey across town. “You should have left earlier, you might easily have missed the whole first act”, your companion reprimands you. Some everyday uses of ‘might’ signal epistemic possibility; but there is, I suggest, an established ordinary language usage which, as in this example, reflects the idea that there are alternative possible ways things might have gone, differing only in relatively trivial, non-fundamental respects – ones which might easily have come about - in which things differ in the respect at issue.) So, I propose to regiment our chosen notions thus: the relevant alternative situations are just those that *might easily* have obtained<sup>6</sup>. Given this regimentation, **Reliability** can alternatively be formulated in the following equivalent condition<sup>7</sup>:

**Reliability** (second formulation): A person who forms a certain belief B by means of a method M she possesses, satisfies **Reliability** with respect to B just if: B is true, and it is not the case that she might easily have formed a relevantly-similar belief B’ by means of M which was false.

In the present discussion I follow Goldman (1979, 1986)<sup>8</sup> and take the belief-forming methods or processes featuring in **Reliability** to be internal to the believing subject – individuated ‘narrowly’, not ‘broadly’.

(Some further comments on details for the precise-minded. First, reliability may be explicated modally, in terms of relevant alternative possible situations, ones that might easily have obtained, as is done here; or it may be explained in actual-world track-record statistical terms - as has been done when explicating reliability that makes for epistemic justification rather than knowledge<sup>9</sup>. These will yield different results in various possible circumstances: a frequently-employed process with a good actual-world track record to date could be modally unreliable in a given situation, or vice versa. I believe the modal conception best explicates our anti-luck intuition concerning knowledge, which is grounded in intuitions about what might easily have occurred, and what could not easily have been different. The modal conception has the advantage that it applies to a process that is in fact employed only once – for instance a single acceptance of testimony. Second, we need to speak of a relevantly similar belief, rather than the very same belief, for several reasons. It is not clear whether that very token belief could have had a different cause; nor is it adequate to say: a belief with the

---

<sup>6</sup> Making this equivalence means one must be prepared to bite this bullet: if which alternatives count as relevant is context-sensitive, then so is talk of what might easily have occurred. I think this could be palatable, but the issues raised by contextualism are beyond the scope of the present enquiry.

<sup>7</sup> ‘Might easily’ as used here is the complement to ‘not easily’ as used in the explanation of a safety condition on knowledge in Sosa (2000). It is a modal notion, not a statistical or probabilistic notion.

<sup>8</sup> Goldman is however here concerned with reliability of processes as making for justification, not knowledge, and accordingly explicates it as a high percentage of true beliefs as output, not our modal conception.

<sup>9</sup> See Goldman (1979, 1986). Reliability making for justification may also be explicated in modally, phrasing the constraint in terms of close worlds; but requiring that most, rather than all, of the close worlds are ones in which relevantly similar beliefs are true.

very same content. Where the proposition believed is a necessary truth it could not be false, so **Reliability**, if applied only to belief in the very same proposition, is trivially satisfied<sup>10</sup>. As to what constitutes relevant similarity of belief, there are murky and crucial issues here that I will skirt around. If we require ‘global’ reliability of the method then we can simply say: not easily would she form a false belief using that same method M. But this is too strong. Consider that I may do an exercise of addition on small numbers and intuitively, my belief is reliably formed. But I may be unreliable in adding together long sequences of very large numbers. So a ‘relevantly similar’ belief needs to be something finer-grained than just: any belief arrived at by using the same method, if we individuate methods coarse-grainedly, for instance ‘addition’.)

What is the epistemological upshot when we apply R-NEC, explicated modally in terms of **Reliability**, to the case of testimony? In applying **Reliability** to testimony I assume first, that the method M is the broad one: formation of belief through acceptance of encountered testimony. (Call a person’s beliefs formed through acceptance of testimony her **T-beliefs**.) I assume that the recipient has a single psychologically real cognitive process of testimony-reception, her (narrow) **T-method**. Her T-method on any occasion will result either in acceptance of and formation of belief in what she is told, or in non-acceptance<sup>11</sup>. And I assume that other ‘relevantly-similar’ beliefs are any other beliefs formed through acceptance of someone’s testimony<sup>12</sup>. This being so, a subject will satisfy **Reliability** with respect to some belief that she forms on an occasion O, through acceptance of testimony, just if it is not the case that her testimonial environment on O is such that *she might easily have encountered testimony (from either the same or another speaker) which is both false, and*

<sup>10</sup> Also, when the proposition is a deeply-entrenched nomological truth that could not easily be false, it could not easily be falsely believed.

<sup>11</sup> Non-acceptance could consist either in outright rejection, or in forming some partial degree of belief short of full acceptance. Note that a person’s total cognitive apparatus in receiving testimony must also involve what we will, for present purposes, regard as a separate cognitive module, call it her **U-method**, her method for understanding what she hears – identifying the force and content of the encountered speech act. Acquisition of reliably true beliefs from encountered testimony requires an effective U-method as well as an effective T-method. The role of successful interpretation of speech acts, and how failures at this point can generate unreliability, is a crucial further topic beyond the scope of this discussion. The supposition of isolation of a person’s U-method and her T-method is an interim simplifying assumption.

<sup>12</sup> If we specify other ‘relevantly similar’ beliefs more narrowly, e.g, confine it to other beliefs acquired from testimony about the same topic, we will get very different results when applying **Reliability** to testimony. E.g suppose my mechanic reliably tells me only truths about my car, but is wildly out on driving directions; I would believe him on any topic, but I ask him about the condition of my car’s engine... Similarly, if we individuate the method as ‘believing testimony given by S’, where S is a particular individual or collection or type of individuals, rather than all potential informants, then modal reliability of this ‘method’ will place different, typically less demanding, requirements. These alternative theoretical choices would alter the significance of the examples discussed in sect.4. I think the relatively broad individuation of both method and other ‘relevantly similar’ beliefs chosen here is apt, but there is not space to probe these issues.

*whose untrustworthiness she would have failed to detect and hence she would have accepted it.* We need to introduce this notion of a **testimonial environment (T-environment)** to explicate **Reliability** in relation to testimony; it is linked with that of what instances of testimony the recipient might easily encounter – these last are just the instances of testimony that the T-environment might yield.

With these points on the interpretation of **Reliability** in regard to testimony settled, we can now investigate the question what R-NEC requires of a recipient of testimony. This is what we find:

**(I) Testimony is Falsity-Prone (TFP):** Testimony as a source of belief is liable to instances of false testimony that are relevant alternatives: it is in the nature of testimony, and is typically the case, that a recipient of testimony in her normal T-environment<sup>13</sup> might easily encounter instances of false testimony.<sup>14</sup>

TFP holds as a law-like general truth about the human social institution of testimony. It is so because human testimony, by its nature, is not an entirely reliable source of belief. False statements made through deceptive intent, or honest error, are an intrinsic and perennial possibility. This being so it is typically the case that a recipient of testimony in her normal testimonial environment might easily encounter instances of false testimony. This being so it follows that:

**(II) Reliability for Testimony Requires an Effective Filter:** A person A who on some occasion O forms a certain belief B through accepting another's testimony satisfies **Reliability** regarding B if and only if she, in her T-environment on O, satisfies **EFFECTIVE FILTER**, where this is the following condition:

**EFFECTIVE FILTER:** A's T-method includes a filter that would detect and block belief in all the instances of false testimony that she might easily encounter; so that she is disposed to believe only the true, knowledge-expressing<sup>15</sup> instances amongst the testimonies she might easily encounter<sup>16</sup>.

**EFFECTIVE FILTER** imposes a substantial condition on a person in her T-environment only if there are indeed some instances of false testimony that she might easily encounter in

<sup>13</sup> Turning on how finely we individuate T-environments, a subject may in fact have a range of normal T-environments she frequents – workplace, home, and so forth. For ease of exposition I prescind from this detail in what follows.

<sup>14</sup> TFP is not the negation of the universal generalization 'testimony is generally reliable'. We have explicated reliability modally, not statistically. So TFP claims nothing about statistics, but rather that instances of testimony, as a type, are liable to have relevant alternatives of false testimony. Elsewhere (Fricker 1994) I have argued that it is a mistake to attempt global generalisations about the reliability of otherwise of testimony, understood in statistical terms.

<sup>15</sup> Fricker 2006 argues that proper instances of testifying are those where the testifier expresses her knowledge; so this condition, rather than mere truth of testimony, is what the discriminating recipient's T-method should be geared to detecting the presence or absence of. I mainly prescind from this detail in the present discussion.

<sup>16</sup> To reiterate: satisfaction of **EFFECTIVE FILTER** is by an ordered pair of a person (by virtue of her narrow T-method), and her T-environment. Her narrow T-method must detect and block belief in all the false testimony which her T-environment might present her with – this being the easily-encountered false testimony.

that T-environment, as TFP insists typically to be so. EFFECTIVE FILTER is trivially fulfilled by a person in her T-environment if the latter is such that there are no instances of false testimony that she might easily encounter. (How the T-environment of a recipient affects how much she herself must contribute in order to secure modal reliability of her T-beliefs is a theme developed in what follows.)

(II) assumes that when someone responds to an instance of testimony her response instantiates a general testimony-processing method, her T-method, which she deploys for all received testimony, and that this is a psychologically real cognitive process internal to her<sup>17</sup>. These presuppositions are in line with Goldman's framework. (In sect.4 we briefly consider relaxing the assumption that the processes relevant for evaluation of reliability are always best individuated narrowly.)

(II) is deliberately stated in broad terms. It insists that the recipient of testimony in her T-environment must somehow possess an effective filter to screen out relevant instances of false testimony, but is noncommittal as to how precisely this is instantiated. EFFECTIVE FILTER is consistent with either a personal-level process that proceeds via the subject's forming reliably true beliefs about the testifier's trustworthiness; or with a purely sub-personal filter, one that may operate without the recipient's attention or any focussed awareness of it. It does insist that this process is *effective*— that is to say, it succeeds in detecting and blocking belief in all instances of easily-encountered false testimony. Only so is **Reliability** ensured.

(II) specifies what is needed for a testimony recipient in her T-environment to satisfy **Reliability** in relation to her reception of testimony. So (II) and R-NEC together imply:

**(III) EF-NEC:** to obtain knowledge from instances of testimony, a recipient in her T-environment must satisfy EFFECTIVE FILTER; that is, her T-method must include a filter that would detect and block belief in all instances of false testimony that she might easily encounter; so that she is disposed to believe only the true, knowledge-expressing instances amongst the testimonies she might easily encounter.

Is TFP compelling? It is - though there are important complications and qualifications, as will be revealed in our subsequent discussion. Testimony is effected by speech acts of telling<sup>18</sup>. As folk linguistics and folk psychology reveal, while such speech acts of telling should, as a matter of social norm, be expressions of the testifier's knowledge (Williamson 2000, Fricker 2006) Fricker (forthcoming) REF, they may fail to be so and state a falsehood if she either is careless and states something she does not know through honest error; or if she

---

<sup>17</sup> That a recipient has a single T-method that she deploys in relation to all instances of received testimony is a substantive assumption that could well be false. It could be, for instance, that she possesses two distinct testimony-processing modules, call them WARY and UNWARY, and that one or the other (but not both) is activated on a given occasion of testimony-reception. The present analysis could cope quite easily with this complication, but I do not attempt to include it in my main discussion.

<sup>18</sup> Extended testimony includes emails, letters, newspapers, the internet, all manner of purportedly factual books and so forth. Here we focus on the core case.

sets out intentionally to deceive, and lies to her intended audience. So it is in the nature of testimony that false testimony is entirely possible, and is a real risk confronting any recipient of testimony. It is a law-like general empirical truth that testimony is not in practice a completely reliable source of belief, if accepted uncritically. This is consistent with great variation in the frequency and manner of false testimony liable to be encountered on different topics, and in different social and cultural settings. This theme will be developed below.

R-NEC and our empirical datum TFP together imply that familiar ‘anti-reductionist’ or ‘fundamentalist’ accounts of what it takes to acquire knowledge through testimony are untenable<sup>19</sup>. They fail to ensure R-NEC is satisfied, since – as we are about to see – they do not impose any condition that ensures that EFFECTIVE FILTER is satisfied. A familiar fundamentalist epistemic principle concerning knowledge from testimony is:

**K-SUFF:** If S testifies that P to H, and in doing so expresses her knowledge that P, and H forms belief in P through accepting S’s testimony, and H possesses no doxastic defeaters<sup>20</sup> for S’s testimony that P, then H thereby comes to know that P.<sup>21</sup>

A recipient of testimony Rita might form true belief through accepting knowledgeable testimony for which she has no doxastic defeaters on an occasion, while her T-Method does not include an effective filter – she would believe false testimony that she might easily have encountered instead<sup>22</sup>. The antecedent of K-SUFF is fulfilled on this occasion, but Rita does not fulfil the modal reliability condition imposed by EF-NEC. If K-SUFF were a correct epistemic principle, Rita would gain knowledge; but R-NEC denies that she gets knowledge, since her belief-forming method is not modally reliable. If modal reliability is indeed needed for knowledge, K-SUFF is too liberal in its apportioning of testimonially-based knowledge, and is not a correct epistemic principle<sup>23</sup>.

The concern of this discussion is with the plausibility of a modal reliability condition on knowledge derived from testimony. We have noted that an account of what it is for a belief to be epistemically justified in terms of its being produced by a reliable process, as pioneered in Goldman (1979), perforce explicates reliability differently, using some precise version of the

<sup>19</sup> The classic statement of and case for fundamentalism is Coady (1990). Wellbourne (1986), Ross (1986), Burge (1993) are further key advocates. For an argument for ‘local’ reductionism see Fricker (1994). See also Lyons (1997), Adler (1994), Graham (2006).

<sup>20</sup> Doxastic defeaters will be beliefs already held by the recipient that show the testimony to be untrustworthy, or which render P extremely improbable on her evidence.

<sup>21</sup> Versions of K-SUFF are advocated by Audi (1997), Adler (2002), Coady (1992), McDowell (1994), Owens (2000), Wellbourne (1986).

<sup>22</sup> That is to say, her T-environment is such that she might easily receive testimony that is false, but for which she possesses no doxastic defeaters, and which is such that her T-method would fail to screen it out as untrustworthy.

<sup>23</sup> Lackey (2008) Ch.2.4 argues against K-SUFF. Her focus is on the fact that K-SUFF does not require doxastic responsibility/rationality of the testimony recipient, rather than on the lack of any provision to ensure modal reliability.

idea: a belief is justified just if it is produced by a process that generates true, as opposed to false, beliefs, a sufficiently high proportion of the time (call this a **suitably veritistic** process). Now a belief can be produced by a suitably veritistic process and yet be false; even when true, it is not guaranteed in virtue of its suitably veritistic genesis to satisfy our modal requirement **Reliability**. Nor is **Reliability** a strictly stronger requirement than suitably veritistic genesis. Nonetheless, it seems likely that suitably veritistic genesis will tend to make for modal reliability, and vice versa – our two conditions, though neither entails the other, may be expected to have a high positive correlation. If this were not so, there would be a theoretically unfortunate divorce between conditions making for justification, and those that are required for knowledge: we would have an alarming amassing of cases of justified true belief that, on our twin accounts of epistemic justification and of a necessary reliability condition for knowledge, do not constitute knowledge although they are not Gettier cases.

In the case of testimony I suggest that a T-method that in the subject's T-environment constitutes an effective filter, thereby ensuring EF-NEC is satisfied, and the subject is equipped to gain knowledge from acceptance of testimony on favourable occasions, is also going to be one that is suitably veritistic, rendering the testimonial beliefs she forms by means of it epistemically justified<sup>24</sup>. That is to say: a T-method that constitutes an effective filter for a testimony recipient in her T-environment will be a normatively apt testimony reception process, one which generates epistemically justified testimony beliefs.

On the other side – to repeat the point made above – it would be unfortunate if our account of conditions for acceptance of encountered testimony to yield justified beliefs came significantly apart from our account of what is required for testimonial knowledge, so that a subject might often acquire justified true testimonial beliefs that were not knowledge, while not being generated through acceptance of a false lemma. This desideratum entails that a fundamentalist account of conditions for acceptance of testimony to yield justified belief are too weak, as follows:

A fundamentalist principle about testimonial justification<sup>25</sup> runs along these lines:

DEFAULT-JUSTIFIED (DEFJUS): If H receives testimony that P, and forms belief that P in acceptance of it, and H has no doxastic defeaters<sup>26</sup> for the testimony, then she is *prima facie* justified in believing that P.

<sup>24</sup> True testimonial belief generated via a suitably veritistic T-method will certainly fail to satisfy **Reliability** when the subject is in a one-off unusual T-environment that is anomalously unfavourable, but may do so more generally since suitable veritism only requires a high percentage of beliefs produced to be true, not all of them. Conversely, in an anomalously favourable environment a belief may satisfy **Reliability** despite not being generated by a suitably veritistic process.

<sup>25</sup> See Burge (1993), Coady (1992), Welbourne (1986).

<sup>26</sup> Doxastic defeaters for an instance of testimony are other beliefs possessed by the subject which are counter-evidence to the truth of the testimony. Possible defeaters for a given instance of testimony

It is obvious that DEFJUS fails to impose any condition on the recipient of testimony H which ensures the modal reliability condition required for knowledge by R-NEC is fulfilled. DEFJUS specifies the absence of doxastic defeaters in a case as a condition needed and sufficing for justified belief in that case; but does not impose on H the general requirement that she would be sensitive to defeaters, if they were present. What precisely would constitute ‘defeaters’ of an instance of testimony remains vague; however it is not so vague that we cannot securely conclude that, even if such a defeaters-sensitivity condition were included, it would fall far short of the **EFFECTIVE FILTER** requirement needed to ensure modal reliability of a recipient in her environment, for many testimonial environments. Consider that a practised liar may not betray any signs of her mendacity. This point is further confirmed in our discussion at the end of sect.2, which underlines how doxastic responsibility does not ensure modal reliability. If it were the case that DEFJUS gives the correct account of what is required for justified testimonial belief, then we would have multiple cases of justified true belief that do not qualify as knowledge but are not Gettier cases – cases where the recipient accepts knowledgeable testimony properly in accordance with DEFJUS, but does not satisfy **EFFECTIVE FILTER**,<sup>27</sup> and which do not involve her belief in a false lemma.

From what we have seen so far it follows that a normative account of what it takes to acquire knowledge through acceptance of testimony which respects the anti-luck intuition of R-NEC will be a ‘reductionist’ epistemology, in that it will impose a requirement that a recipient’s T-method somehow instantiate an effective filter<sup>28</sup>. Note that for R-NEC to be satisfied, the possession of a filter effective in her T-environment is required if a recipient is *on any occasion at all* in that T-environment to obtain knowledge from testimony. To summarise what we have established: Testimony qua epistemic source is liable to relevant-alternative instances of false testimony (TFP). This being so, a recipient of testimony’s T-method satisfies **Reliability**, thereby conforming to R-NEC, if and only if she satisfies **EFFECTIVE FILTER**; that is to say, her T-method includes a filter that, in her T-environment, would detect and block belief in all instances of false testimony that she might easily encounter; so that she is disposed to believe only the true, knowledge-expressing instances amongst the testimonies she might easily encounter.

---

include: evidence that the speaker is likely to be lying, or mistaken – eg. a bad track-record; strong counter-evidence to what she states, including contrary reliable testimony.

<sup>27</sup> A similar rift between justified true belief and knowledge opens up for similar reasons, if a modal reliability condition for perceptual knowledge is combined with a similar account of justified perceptual belief.

<sup>28</sup> Goldberg [REF] maintains that by putting the effective filter at a subdoxastic level a form of anti-reductionism is maintained. I don’t wish to debate the applicability of such labels here. The ‘bottom line’ is: reliability in the process of forming belief through acceptance of testimony must be maintained in part by an effective filter at the recipient’s end.

To put it bluntly: If modal reliability is a necessary condition for knowledge then, if you occasionally do or might easily lie to me, then even when you tell me what you know, I can gain knowledge from what you knowingly tell me, only if I would be able to detect your mendacity, were you to lie to me.

It seems pretty much non-negotiable to accept that testimony is a source of knowledge – that is, that people, members of human societies, are able sometimes to get knowledge at second-hand by means of testimony. This being so, we may formulate:

**Strong Non-Skeptical Constraint (Strong-NSC):** Most of the beliefs that normal people form through their regular everyday acceptance of knowledgeable testimony are in turn knowledge.

Strong-NSC is vague; but not so vague that it fails to impose a very substantial condition that a proposed epistemology of testimony must satisfy, to conform with it. Strong-NSC and EF-NEC together imply:

**EMPIRICAL CLAIM (EC):** Human recipients of testimony in normal society generally employ a T-Method including an effective filter by means of which they would detect and block belief in relevant-alternative false testimonies – ones they might easily encounter.

EC is a substantial empirical proposition; it may or may not be true. If it were established to be false then, given Strong-NSC, *modus tollens* could be applied to show that, after all, EF-NEC and its source R-NEC do not hold of testimony. That is, consideration of testimony would show that modal reliability cannot after all be a global necessary condition for knowledge. In the next section I address the worry that recent research in social psychology does indeed show EC to be false, thus forcing us to either embrace scepticism about testimony as a significant source of justified belief apt to be knowledge, or to abandon R-NEC.

## 2: An Empirical Case Against R-NEC? – The Relativity of Reliability

How ordinary people usually form belief in response to testimony – how critical they seek to be, or are constituted and habituated to be; and whether they succeed in effective monitoring to detect the false testimony they actually encounter or easily might encounter, in the social environments they typically find themselves in – is an empirical question. Thus, whether current human testimony-reception practices and habits conform to proposed necessary conditions for knowledge developed from the armchair by philosophers is not a question I shall answer definitely in this discussion.

Joseph Shieber is a philosopher who has attended closely to empirical studies of testimony-reception; and he has drawn negative conclusions about the conformity of ordinary people with the normative requirements suggested by philosophers (Shieber 2011). In this

section I will suggest that his pessimism is not warranted by the evidence from social psychology that he cites. The empirical studies he cites do not establish that ordinary people do not and are unable to monitor effectively for truthfulness of the testimony they encounter in their social transactions in their everyday environments: they do not show that our empirical claim EC is false.

TFP records the indisputable fact, certified as true by our folk understanding of the nature of the speech act of telling and the human failings it is subject to, that testimony qua epistemic source is liable to false instances – either through deceptive intent, or through honest error. But the global truth of TFP is consistent with great cultural variation in the frequency and nature of false testimony that recipients of testimony are at risk of encountering, in their normal social environment. In addition, within a given broad cultural-social environment, there may be very specific local variations between contexts in the likelihood and character of possible false testimony that recipients run the risk of encountering. Reliability of a process in producing a given outcome is, as a matter of logic, relative to a range of possible inputs to it, and so to an environment supplying these inputs in which said process operates. What it takes to be an aptly discerning recipient of testimony – one who effectively detects and blocks belief in the instances of false testimony actually or easily encountered – varies, is more or less demanding, according to that recipient's testimonial environment: what frequency and what manner of false testimony she might easily encounter.

Suppose a research team receives funding for an investigation of the reliability of ordinary people's visual perceptual judgements. Suppose they set up an experiment to test this as follows: they construct an environment, through which the experimental subjects are conducted and are asked to report on what they can see around them. This environment contains a series of perceptual illusions and tricks – holograms, trompe l'oeil paintings, trick mirrors, and so forth. Suppose first the subjects have not been told of the tricky nature of their environment. Naturally, their perceptually-based beliefs, reported from their tour, display a considerable degree of falsity – they were tricked by the trickery. But even in the variant experiment where they are forewarned, this will not suffice to eliminate falsity in the beliefs they form; unless a subject adopts the desperate strategy of forming no beliefs from what she seems to see at all. The team writes up their research and concludes that visual perception is a highly error-prone and unreliable source of belief. This conclusion is not warranted by their findings. Or rather: their findings are correct understood as concerning only the subjects' performance in the environment in which the experiment placed them. It is deeply misleading, if interpreted as having significance for the general reliability of visual perception as a source of belief outside that specific tricky environment. Reliability of a

process or faculty is, as a matter of logic, relative to an environment in which it operates. So we must beware of exporting findings about reliability in a particular environment or situation beyond its specifics, to other situations that may be very different in relevant respects.

Consider now the idea of a perceptually-based recognitional capacity. A creature has a capacity to perceptually recognise some type of item *iggs* just if she can perceptually discriminate *iggs* from the relevant-alternative non-*iggs* – that is, all those items that are not *iggs* and that she might easily encounter. (Anything at all which is not an *igg* is a non-*igg*; but we are concerned with those non-*iggs* which a subject might have some trouble in telling apart from *iggs*.) Discriminations in humans will issue, amongst other responses, in perceptually based judgements – ‘That’s an *igg*’/ ‘That’s not an *igg*’. In other creatures they may trigger forms of behaviour apt when in the presence/absence of an *igg*.

Now consider this scenario: I am a keen observer of the birds that come into my garden. There are various species members of which show up regularly, others occasionally, yet others, never. In this situation, at home looking out of the window into my garden, I have a recognitional capacity for robins. I am disposed to form beliefs such as: that is a robin/ that is not a robin, on the basis of the visual appearance to me of any bird I spot in the garden, and these beliefs are formed by a modally reliable method, since I can indeed recognise a robin by its appearance. That is to say, I can discriminate robins by how they perceptually appear to me, in contrast with the relevant alternatives – the other bird-species members of which from time to time show up in my garden. Relative to the environment of my garden, the inputs it provides, my perceptual method for forming beliefs about the presence or absence of robins is modally reliable.

But then I go on holiday, to a distant place where various other bird-species also show up regularly. These include bobbins – birds of a different species, with very different habits but which, while being slightly larger and heavier, have very similar colouring, shape and demeanour to robins. I fail to discriminate between robins and bobbins, mistaking bobbins for robins. Even when I learn about the existence of bobbins, I cannot tell the two species apart on the basis of how they appear to me.

In this new context I lose my homely recognitional capacity for robins. Any beliefs I form about their presence or absence on the basis of visual perception, even where true, will fail a modal reliability condition. When I stayed safely at home I could tell on the basis of perception whether or not a bird I saw was a robin. In my new holiday environment I lack this ability. In this changed environment my very same ‘narrow’ perceptual processing method fails on the task – recognising robins - it successfully fulfilled back home. I have not changed, but my environment has changed, introducing new relevant alternatives.

The logic is exactly similar when we consider a recipient’s ability to tell when she is being lied to. Ella, a teenager, has a circle of friends in which there is a strong norm of trust

and honesty. They very rarely deceive each other – even to the point of preferring honesty to tactfulness. In the situation of Ella and her circle lying is infrequent, and anyone who lies is unpractised, guilty and embarrassed. The liar shows tell-tale signs and Ella, no fool and perceptually well-attuned, is able to detect them. Then Ella's father gets a new job, and the family moves to a city in another part of the country. Ella goes to a new school with very different social conditions and mores. In her new school, lots of the pupils lie regularly in pursuit of their own selfish purposes and are practised and proficient dissemblers. Ella retains the perceptual attunedness she previously possessed – her 'narrow' T-method is the same. But in her new environment the old skill is not sufficient for the different and more taxing task of detecting when these cynical streetwise city kids, her new classmates, are lying.

The moral of our two tales is evident: possession of a discriminatory capacity for iggs enabling satisfaction of a modal reliability condition for judgements issuing from it is relative to an environment. The environment fixes the set of relevant alternative non-iggs that the perceiver must be able to discriminate against. A perceiver who possesses a discriminatory capacity in one environment will lack it in another environment, without any change in her 'narrow' capacities, and due entirely to the difference in her environment.

It is an instance of this broader phenomenon that whether a recipient possesses an effective filter to block instances of false testimony, ones that she might easily encounter, is only determinate relative to an environment. We saw this in the tale of Ella, and the point generalises.

To summarise: EFFECTIVE FILTER specifies that a testimony-recipient's T-method includes a filter effective for the instances of false testimony that she, in her T-environment, might easily encounter. But we have seen that whether this is satisfied is the upshot of the fit of the recipient's 'narrow' T-method with her external social environment - her T-environment. EFFECTIVE FILTER is trivially satisfied by a recipient when her T-environment affords no instances of false testimony that she might easily encounter. More generally, how powerful a recipient's internal filter needs to be – how finely-attuned her capacity to detect falsity of an instance of testimony - turns on what the range of false testimony she might easily encounter is like. As we saw with Ella: if a recipient finds herself in a community where lying is rare, and liars tend to be inept, revealing their mendacity by tell-tale signs in their demeanour, then it is relatively easy for her to possess a T-method with a filter effective in that T-environment. If, in contrast, she finds herself in an environment of habitual and expert liars, the cognitive requirements for possession of a filter effective in that

T-environment are much more taxing, maybe impossible to fulfil. (Much more on this relation between filter and environment in my final section.)<sup>29</sup>

The environment-relativity of discriminatory capacities has immediate bearing on our present topic, whether the empirical findings cited by Shieber show that EC is false. The negative empirical findings, assuming their methodology is sound,<sup>30</sup> show that subjects are unable to achieve any high percentage of success in their attempts to detect on the basis of their ‘demeanour’ (i.e. from cues in their verbal and non-verbal behaviour) the ‘liars’, that is falsity-speakers<sup>31</sup>, amongst a series of separately-observed individuals; each of whom reports for a very short time on a particular topic, in a very specific experimental set-up. (For surveys of the experimental data see (Kraut 1980, Zuckerman, DePaulo et al. 1981, DePaulo 1985, Ekman and O’Sullivan 1991, Bond and DePaulo 2006). Even if the soundness of these results is accepted, to infer from them to the truth of the sweeping general claim that ordinary people do not monitor effectively for truthfulness of the testimony they encounter in their everyday environments (the denial of EC), requires the additional premiss that the T-environments of the experiments, and those of T-recipients in their everyday transactions, are similar in terms of what challenges – what manner of easily-encountered false testimony – they threaten. Nothing in the experimental data itself supports this premiss, and first-off it seems more likely to be false. In any case the experimental findings are mixed, rather than consistently negative. (Ekman, O’Sullivan et al. 1999) finds individual variation, some individuals doing quite well at detecting lying, others less well. (Zuckerman, Koestner et al. 1984) finds improvement in detection of lying of particular individuals by experimental subjects, when given feedback on their success/failure.

I have suggested that the empirical findings cited by Shieber, since they concern recipients’ lack of effective filtering capacities in certain very specific testimonial situations, do not demonstrate, or even make probable, the thesis that ordinary recipients of testimony do not possess an effective filter which allows them to satisfy the modal reliability condition of R-NEC, in their typical everyday social environments. A worry may be raised here thus: “True, the data which show that recipients of testimony are bad at detecting lying concern

---

<sup>29</sup> Flukes apart, someone’s testimony is true just if they are both sincere (express their own belief in their statement), and the belief they express is true. The empirical literature focuses on detection of lying, rather than of honest error, and so I do so for much of this discussion.

<sup>30</sup> I am not trained as a social scientist, fully to critique the soundness of experimental methodology and validity and robustness of results. A full evaluation will need to do this.

<sup>31</sup> I insert this caveat because it is of course extremely difficult in an experimental situation to recreate the real situation of a true lie – someone who is deceiving with bad faith, for her own gain, due to her own motives, and with attendant fear of discovery and feelings of guilt. College students being paid to misdescribe their feelings about what they are seeing on a film is a good attempt, but falls short of the real thing. See Ekman, P. and M. O’Sullivan (1991). “Who Can Catch a Liar?” *American Psychologist* 46(9): 913-920.

studies in a very specific experimental setting; but the nature of the findings may nonetheless be such that it is likely that they will carry over to other situations in which testimony is given and received.” One reason to doubt this, is that liars are not all alike. They all share a deceptive intent; but they will be more or less practised in the art of effective deception – as our school-girl toy example above illustrated. Supporting the worry are the empirical studies which find that the features that are in fact tell-tale signs of lying are different from those which folk-theory about lying believes to be key. In particular the studies indicate that sustained eye-contact is not, as folk wisdom has it, a sign of truth telling (Stiff 1989).

There is an important general reason why the results from the studies are of only marginal relevance to the truth or otherwise of EC. The studies are explicitly aimed at investigating how good subjects are at detecting when a speaker is lying from cues observable in her demeanour. This is their topic. But when we turn to whether a T-recipient has an effective filter against false testimony she may encounter in her everyday environment, her method is not confined to these limited materials. An effective T-method for detecting, or suspecting, lying, will for starters draw on the T-recipient’s background knowledge – both general social knowledge, and specifics of the testifier in question. A discriminating, non-gullible T-recipient will approach an instance of testimony in the light of her empirically-based assessment of whether this type of person is generally truthful about this type of subject; whether this particular testifier has a past record of mendacity; and, most crucially, whether a plausible interpretation of her purposes and projects in speaking gives her a *motive* to lie. This is a wealth of relevant information available as input to assessment of likely mendacity, quite independent of whatever tell-tale signs may or may not be present in the testifier’s demeanour. If awareness of these factors – very possibly achieved without conscious attention – is enough to cause the non-gullible recipient to take the offered testimony ‘with a pinch of salt’ – that is, not to simply believe it outright; then this is enough to avoid gullibility, and failure to satisfy **Reliability**. Cues in behaviour are not the be-and-end-all, in assessing for likely mendacity; and options for doxastic response are not confined to: accept outright, or reject entirely. This I think shows up one way in which the typical purpose of the empirical work makes it not directly relevant to the issue of the modal reliability, in everyday situations, of recipients’ T-methods. The study (Ekman and O’Sullivan 1991) of whether police interrogators and other similar professionals can tell on the basis of perceptual cues which out of a series of otherwise very similar speakers on the same topic are in fact lying, looks for the presence of a very fine-grained discriminatory capacity. This is not the same issue at all, as the one sketched above – whether and how ordinary recipients of testimony, in ordinary social situations, can engage in relatively coarse-grained, but mostly effective, sifting of types of testifier by social role and other social-background cues, together with

deployment of everyday folk-psychological know-how to estimate the speaker's likely purposes in speaking, and thence whether she has a motive to lie.

If it were the case that one gains knowledge through perception only if one's perceptually-based beliefs are modally reliable; and if the test for this were whether one would be able to tell, if one were being deceived by an incredibly powerful demon intent on deceiving one; then, inevitably one would not be modally reliable, hence not knowledgeable, in believing anything much more than that one exists. But, as is familiar, this supremely high standard is not apt, except perhaps in certain very specific contexts, for assessing the reliability of perception. Equally: if one gains knowledge through accepting testimony only if one's T-method is modally reliable; and if the test for this is whether one would be gulled into acceptance by a *perfect liar*, one whose motives one had no antecedent reason to suspect, then inevitably one would not be justified in believing anything on the basis of another's testimony. But, as with perception, this supremely demanding test is not the epistemically apt one to invoke, as a standard for knowledge. Nor is it the usual, and correct, interpretation of R-NEC.

I have focussed so far on the detection of lying, since this is the focus of the empirical studies. But competence or its lack – being knowledgeable about her topic – is no less important a factor than sincerity in the overall trustworthiness of a testifier. We all are often more at risk of encountering testimony that is false due to incompetence, than to mendacity. Shieber cites experimental findings concerning this also. Here, I think the findings give grounds as much for optimism as for pessimism, about the effectiveness of ordinary T-recipients' current strategies and capacities. Several studies found that, in assessing informants for competence, recipients are sensitive not (*per impossibile!*) to competence itself, but to the social trappings and concomitants of competence. Thus, people presenting themselves as doctors, and appropriately attired, are believed on medical topics more than those who present as inexperienced laypersons; those in a guard's uniform are obeyed more readily than those in work-clothes (Bickman 1974). This strategy, though in epistemically (and socially!) unfavourable situations it would not lead to detecting incompetence, is an apt implementable strategy, in socially and epistemically favourable situations. - Who should one believe as to what is wrong with one - the doctor, or one's neighbour? Whose instructions should one follow, as to what to do when fired shots are heard in an airport – the security guard, or another traveller? It is simply an instance of the inevitable general strategy in non-omniscient knowledge-seeking beings, for gaining knowledge about non-immediately perceptible features of the world. A creature in a forest seeking food eats plants with a certain appearance generally correlated with wholesomeness, and flees from other creatures similar

in appearance to dangerous predators. In the supermarket, I look for marmalade, by looking for jars of the appropriate size and appearance with labels saying ‘marmalade’. Similarly, and with no less epistemic aptitude, I trust the man at the garage in mechanic’s overalls, oily and holding a spanner, about what is wrong with my car - and so forth. While not an infallible procedure – and, *ex hypothesi*, not one that will be effective in identifying professional incompetence – this is an epistemically effective strategy, in a moderately epistemically benevolent and socially well-ordered world: one in which my doctor has diagnosed my common complaint successfully, and my confident but inexperienced and paranoid neighbour has not. It is not an effective strategy in an epistemically and medically disastrous world, where medical schools turn out money-grabbing and incompetent professionals. (We will return to the significance of this point in sect.4.) But – as I have sought to highlight - a T-recipient needs a T-Method with a filter effective in the epistemic and social environment she finds herself in; not one proof against socio-epistemic demon-worlds.

We have seen that the experimental data cited by Shieber do not relate directly to the typical epistemic predicament of ordinary recipients of testimony on commonplace everyday topics. We have seen also that how powerful a filter a T-recipient’s T-method must incorporate, to be effective, is a function of her socio-epistemic environment.<sup>32</sup> This suggests that our original empirical claim EC is too sweeping to be a good candidate for a law-like generalisation, whether true or false. Whether individual T-recipients in a particular culture, society or more specific local social context mostly satisfy EFFECTIVE FILTER will turn on the nature of that culture, society or specific social context. So an empirical claim on the lines of EC is best formulated as an empirical proposition about such a particular culture, society or context. And, I conjecture, some of these context-relativized versions of EC will be true, others false – turning on the contingent and variable cultural norms and practices surrounding the production and reception of testimony across these contexts. In a cultural environment that is epistemically favourable, that is to say false testimony is rare and bears easily recognisable signs, then EC is likely to be satisfied with respect to it. In one where hard-to-detect false testimony abounds, or where the socially-reinforced epistemic norms governing acceptance are very uncritical, it will very likely be false.

I have argued that the studies Shieber cites do not provide premisses for a safe inference to the conclusion that EC – better, many of its many relativisations - is false. But EC may yet

---

<sup>32</sup> See also Michaelian (2009). Michaelian takes the empirical literature to have established that speaker ‘monitoring’ for truthfulness is ineffective. He points out that ineffective monitoring for deceptiveness, and indiscriminating acceptance, serve equally badly as T-methods in a T-environment that offers a high proportion of (*ex hypothesi* undetected) liars; but also equally well in one offering very few liars. However he does not address the issue of the limited ‘ecological validity’ of the empirical studies which has been my focus here; in particular the fact that a situation in which screening for lying is confined to perceptual cues as its basis is not a T-recipient’s normal situation.

be predominantly false nonetheless. How much trouble would this engender, for R-NEC? My next section considers this question.

### **3 Finessing the Non-Skeptical Constraint**

We saw that R-NEC together with Strong-NSC entails the empirical claim EC: ‘Human recipients of testimony in normal society generally employ a T-Method including an effective filter by means of which they would detect and block belief in relevant-alternative false testimonies – ones they might easily encounter.’ In the previous section we saw that the evidence from the studies cited by Shieber does not demonstrate EC to be false, or even render this probable.

This does not, of course, show that EC is true; our interim conclusion is that the matter remains open to further enquiry.

But suppose that further empirical studies were securely to establish the following: ordinary recipients of testimony are quite often in environments where they, as things stand, do not employ a T-method that provides an effective filter in that environment – that is, there are currently many actual socio-epistemic environments, and T-recipients embedded in them, relative to which EC is false. This would mean that, unless we give up on R-NEC, Strong-NSC (‘most of the beliefs that normal people form through their regular everyday acceptance of testimony are knowledge’) would not be satisfied. Would this be enough to precipitate a crisis for reliabilist epistemology, suggesting R-NEC must go? I shall now explain why it would not do so.

I suggested in sect.1 that it is non-negotiable to accept that testimony is a source of knowledge– that is, that people, members of human societies, are able sometimes to get knowledge at second-hand by means of testimony. But this compelling thought does not entail Strong-NSC, but only a weaker non-skeptical constraint:

**Weak-NSC: Testimony for humans potentially can be, and sometimes is, a source of knowledge.**

Strong-NSC portrays current everyday habits and practices of testimony-reception as epistemically apt, furnishing belief apt to be knowledge, just as they currently are. But Weak-NSC does not entail that we must endorse current social norms and attitudes to testimony,

including recipients' receptive practices, as epistemically ideal just as they are, with no scope for improvement. Epistemology is of small interest if it is not normative; and its proposed norms are of small interest if we hold a meta-epistemology which entails that all the philosopher can ever do is to slavishly record and put a seal of approval on present-day epistemic habits and practices. To endorse Weak-NSC as compelling is one thing. We should not move from there to the feeble view that one cannot challenge and criticise current everyday epistemic habits and practices; and that our normative epistemology of testimony is constrained to come up with the result that all or anyway most instances of actual testimony-reception are normatively apt, and apt to result in knowledge. Strong-NSC, while it may possibly turn out to be satisfied *ex post*, is not a binding *ex-ante* constraint on an adequate normative epistemology of testimony. The upshot of the confrontation of armchair normative theorising with the results of empirical research about typical present-day epistemic habits, including testimony-reception, could turn out to be a critique that suggests present habits are far from epistemically perfect. For all we yet know (and as we may well suspect!) socially-reinforced and maintained epistemic habits and culture in relation to testimony are epistemically defective both as regards modal reliability, and as regards doxastic responsibility. Relative to the epistemic goal of attaining true belief while avoiding error, we may find ourselves concluding, habits need to change. Testimony-recipients may need to be warned, informed, educated, to do better than they currently do. Specifically, we might find that many recipients are needlessly gullible, open to being deceived, and that – relative to the epistemic goal of attaining true belief while avoiding error - they would do well to learn to be more suspicious and discriminating, and more circumspect<sup>33</sup>.

Weak-NSC is however compelling. This being so, it *would* precipitate a crisis for normative armchair epistemology of testimony, and specifically for R-NEC, if it were shown both that:

- (a) It is a robust and irremediable empirical feature of human testimony, that there is lots of lying and/or incompetence across nearly all societies; and also that
- (b) Humans are constitutionally incapable of learning to respond discriminatingly to testimony.

If both (a) and (b) were empirically established, the catastrophic conclusion – catastrophic for the human race no less than for reliabilist-minded epistemologists - would follow that the

---

<sup>33</sup> I am of course not advocating a general programme of compulsory epistemic re-training, or other unacceptable forms of social engineering. But first, information may move people; second, for specific professionals it may be acceptable to require training in appropriate epistemic practices; third, education already involves some teaching of how to be critical of sources, thinking about possible bias, angles and so forth. More of this could empower citizens by enabling them to seek information more effectively from available sources – which websites to trust, how to detect bad information, and so forth. A different topic from today's; but one suggested by it.

empirical facts preclude EFFECTIVE FILTER and hence **Reliability** being satisfied, in relation to any instances of testimony. We would then face a choice between dropping R-NEC – to allow that knowledge is gained from some instances of true testimony that are believed; or espousing scepticism about testimony as a source of any knowledge at all. Fortunately the catastrophic conclusion is not even remotely probable. It is hard to see how (b) could be empirically proven. True, some empirical studies show people tend to be overconfident of their own abilities, and to respond badly to critique of them<sup>34</sup>. But the human animal is supremely the learning animal. It is hard not to think that humans are, at the very worst, capable of learning, through confrontation with conclusive evidence, that they are bad at telling who is lying in a given situation, so that they should suspend judgement on this matter; and acting on this information. This is all that it takes to prevent the formation of modally unreliable belief, in violation of R-NEC<sup>35</sup>.

We saw in 2 that the cognitive demands on a recipient, for her to embody an effective filter are more or less taxing, according to what manner of false testimony might easily be encountered by her; EFFECTIVE FILTER being trivially fulfilled in a T-environment in which there is no false testimony easily encountered. This suggests evading the catastrophic conclusion by denying its other premiss, (a). The nature of the socio-epistemic environment, in particular social and institutional norms governing the giving of testimony, and sanctioning defective testimony, is not an unalterable given. Even if human nature cannot be changed, human social arrangements are eminently subject to flux and alteration, in this case with benign intention. This thought will be pursued further in my final section.

The considerations of this section underwrite and strengthen the case for the conclusion we came to in 2: there is no crisis for R-NEC precipitated by the empirical studies of human testimony-reception cited by Shieber. In 2 we saw that EC is not disconfirmed by them. In this section we have seen that even if EC were falsified, this would not force the abandonment of R-NEC, since only Weak-NSC is a compelling ex ante constraint on an adequate normative epistemology of testimony. It would not be out-of-bounds if the confrontation of normative epistemology with empirical data were to yield the conclusion that there is room for improvement, from the epistemic standpoint, in current practices as to both the production and reception of testimony.

We have found that the need to avoid an unacceptable degree of scepticism about testimony as a source of knowledge does not force the abandonment of R-NEC in relation to testimony. But this does not show that modal reliability is in fact an apt necessary condition

---

<sup>34</sup> See Dawes (1994), and also the discussion and studies cited in Kornblith (2002).

<sup>35</sup> Suspending judgement in a case of possible lying would not directly yield a true belief, of course; but it could ensure the subject had an effective filter, and so enable other instances where she accepts testimony to qualify as knowledge.

for testimonial knowledge. In sect.4 I develop a more classic apriori line of thought, presenting a series of cases which I shall attempt to persuade you make the requirement of modal reliability of a belief, for it to be knowledge, seem over-strong in relation to testimony. Before this, I make a brief digression to contrast R-NEC/EFFECTIVE FILTER with internalist accounts of justified testimonial belief that trade in doxastic responsibility rather than modal reliability.

A normative epistemology of testimony may place lesser or greater cognitive demands on a recipient, as required for a belief she forms through accepting testimony to be justified, and hence potentially knowledge<sup>36</sup>. Label as **Blind Trust** the T-method of a recipient who **uncritically accepts and forms belief in whatever she is told, whomsoever by**. Unqualified Blind Trust is not a coherent stance – since one may simultaneously receive inconsistent testimonies from different people. To preserve minimal coherence in the blind truster's cognitive economy, she must be sensitive to the most blatant defeaters of the received testimony. As well as contrary testimony, these include being told by another source that the testifier is lying; or already having conclusive evidence, such as excellent perception, of the falsity of what is testified to. Suppose Blind Trust to be refined to cope with these points. So far as I am aware, no writer on testimony has subscribed to the view that Blind Trust is an epistemically apt attitude. The fundamentalist positions we examined in sect.1, K-SUFF and DEFJUS, both require sensitivity to defeaters, which we may interpret as extending beyond the bare minimum required to avoid glaring incoherence. Further defeaters of the supposed presumptive justification to accept as true what one is told include various obvious warning signs that the testimony is untrustworthy: grounds to think the person has a motive to lie; a previous track record of bad faith or inaccuracy; knowing they are not in a position to know about the topic; and so forth.

Epistemologically internalist approaches place requirements for justified belief formation through testimony designed to ensure that the belief in question satisfies conditions of doxastic responsibility/epistemic rationality. A subject forms a belief responsibly/rationally if and only if she has grounds or reason to think it likely to be true, from her own epistemic perspective. Internalist **fundamentalists** hold that it is enough for this condition to hold, that the recipient has no defeaters for the testimony. They maintain that testimony has intrinsic credibility, and so justifies acceptance by its own light, so long as this is not defeated. Internalist **reductionists** instead hold that the word of another is not intrinsically credible, so

---

<sup>36</sup> I take justification to be a necessary condition for knowledge; if one's favoured reliability condition does not ensure this, then it is a further necessary condition.

that one rationally accepts a proposition on the basis of testimony to it only if one has specific positive empirical grounds to trust that testifier about her present topic.<sup>37</sup>

We saw in sect.1 that fundamentalist conditions for justified belief through testimony do not ensure that a subject forming belief in accordance with them satisfies a modal reliability condition. A testifier could give true, knowledgeable testimony, and a recipient, having no defeaters for it, form belief in it, properly according to DEFJUS; while in fact the testifier might easily have given false testimony, which the recipient would equally have believed. The same holds for the reductionist internalist requirement. A recipient could have excellent positive reasons to trust a piece of testimony, and nothing against doing so; while in fact the testifier, though speaking truly from knowledge, might easily have lied undetectably about the same or a similar topic.

There is no question of getting knowledge that P from testimony that P, when P is false. But when the testimony is true, and the recipient has excellent positive reasons to believe the testimony, and none against – does it really prevent her coming to know that P, simply because the testifier easily might have told her something false instead, and the recipient would have believed her? R-NEC gives an affirmative answer to this question. In these circumstances the recipient's belief, though entirely epistemically rational, and true in virtue of its source in the testimony, is not modally reliable. In the next section I will question whether modal reliability as formulated in **Reliability** is really such a good idea, as a necessary condition on testimonially-based belief, for it to be knowledge.

#### **4. Justification without Reliability?**

Goldman (1976) theorises knowledge as belief formed by a modally reliable process or method, such methods being individuated narrowly. EFFECTIVE FILTER formulates what this requires of a recipient of testimony. We saw in 2. that how powerful a subject's internal filter (her narrow T-method) needs to be depends on how epistemically favourable or otherwise her T-environment is – what manner of false testimony she might easily encounter, and so must be able to detect. At one extreme, in a testimony demon-world, filled with randomly distributed perfect liars and incompetents, even the most powerful possible narrow T-method would not be able to ensure modal reliability. At the other, Blind Trust is a modally reliable T-Method, in an environment where one encounters only true testimony. Equally, one whose narrow T-Method is Blind Trust will satisfy modal reliability if she is everywhere accompanied by a Guardian Angel, who intervenes to convince her of the contrary truth or block belief, if any false testimony is offered.

---

<sup>37</sup>See Fricker (1994), Graham (2006)

Blind Trust does not seem a very good T-Method, even for someone who has a Guardian Angel. To reveal what is epistemically weak about Blind Trust we can evaluate (narrow) T-Methods not just in terms of their modal reliability in the recipient's actual environment, but across a range of possible environments.<sup>38</sup> Blind Trust does badly once a range of Guardianless, and less-than-ideally favourable testimony-environments are included. We may say that a T-Method is **robustly modally reliable** if it has modal reliability across a good range of possible T-environments.

A T-recipient with a robustly reliable T-method is better equipped epistemically than one with a merely actually-reliable T-Method. (Think of Ella's move of school, and my move to a new bird-watching location.) A T-recipient's T-Method will be the more robustly reliable, the more the burden of ensuring reliability is taken by apt sensitivity to cues (both perceptible features of the testimony, and relevant background information) in her internal filter, rather than being due to contingent beneficial features of her environment. If I would succeed in blocking belief in the false testimony I might easily receive, only in virtue of the fact that a constant companion to whom I defer would interrupt and advise me of the fact, whenever I am given false testimony, this is not a robustly reliable T-Method – take away my companion, and I am easily duped. The more powerful my internal filter, the greater the range of increasingly unfavourable T-environments in which my T-Method will still satisfy modal reliability.

There could equally be negative environmental interference messing up the outputs of an otherwise quite robustly reliable internal filter. Consider the converse of the Guardian Angel set-up: I discriminate false testimony on the basis of perceptual cues. But when I encounter, and detect, a false testimony, my Host Demon interrupts and provides me with misleading but persuasive defeaters for the falsity of the testimony.

The Guardian Angel and Host Demon count as features of my variable environment, not as part of my T-Method, given our narrow individuation of methods. But if an outside-my-skin device is bound always to be with me, one could instead individuate my T-Method to include this external device. This is surely a good theory choice, for instance, for the visual-perception method of short-sighted people who always wear glasses. It could be so also, for someone who always receives testimony along with her effective lie-detector machine<sup>39</sup>.

---

<sup>38</sup> See the discussion of related matters in Goldman (1999) ch.4.

<sup>39</sup> Goldberg (2008) suggests that to give a plausible account of how young children are able to get knowledge from testimony despite being unable to filter effectively, we need to theorise their T-method as including their carers who in effect, as he maintains, serve as Guardian Angels. Goldberg (2010) goes further, proposing a more radical externalism in the individuation of testimony as a belief-forming process, to include the testifier as part of a single 'broad' cognitive process. Goldberg here echoes the strategy of Burge (1993), treating testimony as metaphysically and epistemically analogous to memory – a single 'rational source for reason' within the cognitive economy of a rational subject. It is beyond the agenda of the present discussion to tackle this proposal, but I am as yet unconvinced.

A robustly reliable T-filter is more powerful epistemic equipment than a solely actual-world reliable filter. Our present project is to question whether even actual-world reliability is needed in all testimony cases which intuitively count as knowledge. Consider the case of Cyril and Maud. Cyril's spouse Maud is having an affair. Occasionally she comes home late having been with her lover. More often she comes home late after staying late at the office to finish urgent work. In either case she tells Cyril that she has stayed late at the office to work. She has always been an exemplary spouse until this recent affair, and Cyril has lots of track-record evidence of her past honesty and competence. He has no reason to suspect anything currently. Maud is a very adept liar, good at dissembling. Cyril cannot tell from her demeanour when she is lying, and harbours suspicions neither of her mendacity nor her unfaithfulness. All this being so, Cyril is entirely doxastically responsible and evidentially sensitive in his acceptance of his wife's reports of where she has been. He satisfies even a strong reductionist internalist requirement of possessing positive grounds to treat Maud's testimony as truthful. Of course Cyril does not get knowledge of where Maud has been when she lies to him and he believes her false report. But what of the good case, when Maud sincerely and truly tells him that she has been working late? His belief is true and doxastically responsible; but it fails our modal reliability condition – she might easily have lied, and he would still have believed her. My intuition is that in the good case, when Maud is telling the truth, Cyril comes to know through her testimony that she was working late at the office. If this is right, then our usual modal reliability condition on knowledge **Reliability** is incorrect, at least for testimony.

(Of course, if Cyril comes to know that Maud is having an affair, and sometimes lies to him about where she has been, that additional knowledge will undermine his trust in her testimony, even when she is telling the truth. Because this defeater has come within his ken, the internal doxastic responsibility condition now kicks in, and if rational he will no longer be entirely confident in trusting her word. Even if he were still to trust her this would be epistemically irrational, and so his belief would not be knowledge. Such loss of knowledge through coming to know misleading factual defeaters is a familiar phenomenon, and is not a reason to deny knowledge when the believer remains ignorant of the misleading factual defeaters.)

We can play around with details here, and intuitions maybe shift a bit. Suppose Maud is nearly always with her lover, and by small chance this is the one occasion when she was at the office, and speaks truly. Or suppose she tells Cyril truly 'I was at the cinema with a friend from work.' This is true – her lover is from work, and we may suppose counts as a friend. But Cyril might easily have asked who the friend was, in which case she would have lied and said it was her colleague Jane - relevantly similar false testimony. My intuition is that Cyril gets knowledge from acceptance of Maud's testimony in all of these cases. But I think it is

especially hard to deny it, in our first and third cases.<sup>40</sup>

Why do we think this? (I address those who share my intuition, and invite others to join us!) Consider the following scenarios. I need urgently to get my prescription from the chemist. I by fluke get to the chemist just before it closes, but only because (unknown to me) today is late-closing day. Or: I go by chance to the only chemist that has the prescription; all the others in town do not stock it. In these cases I get the goods I need not through my own powers, but only due to lucky chance. It could easily have been that I failed to get them. But I did get the goods nonetheless!

Some knowledge-getting cases are uncontroversially like this. By chance, I catch a glimpse of the burglar as he leaves, and recognise him. So I know it was Tom Farley. If I'd not seen him, I would have falsely believed it was Jonas Perkins. This case is of course not a counter-example to R-NEC; **Reliability** is satisfied. My belief is not brute-modally-reliable, but it is so relative to my method: visually-based perceptual recognition. I got the goods, I got knowledge, though only through a chance event; only through chance did I have access to the epistemic link that enabled me to gain the knowledge.

Testimony, I suggest, furnishes cases – those structurally similar to Cyril and Maud – where we have the same intuition – I got the goods, though only by chance. But in these testimony cases my belief does not satisfy **Reliability**. The belief is not modally reliable even relative to its method<sup>41</sup>.

There are several kinds of cases we can think up, concerning testimony, where (to start with an uncontroversial description) the subject gets a true belief from acceptance of testimony, but one which does not satisfy **Reliability** – she might easily have obtained a relevantly-similar false belief, through acceptance of other relevantly-similar testimony. They are all cases where **Reliability** is failed; but I think our intuitions discriminate between them. If I am right about this, then it is not just whether a modal reliability condition is failed that matters, but how and why it is failed.

---

**Reliability-Failure Case One:** I receive **accidentally true testimony that P from a speaker who lacks knowledge, and has a true belief regarding P only by chance**<sup>42</sup>. This speaker might easily have had a false belief about this topic, and would have misinformed me.

---

<sup>40</sup> Compare the case when Maud says only 'I stayed late at the office', letting Cyril think she was working, when in fact she and lover were together there. Fricker (2012) explores the respects in which misleading someone by means of conversational implicature differs from explicit lying.

<sup>41</sup> Assuming, that is, we do not do anything fancy with our individuation of methods to preclude the negative result. See below.

<sup>42</sup> Note that this is not a Lackey-case, viz. of a reliable testifier who nonetheless lacks knowledge. See Lackey (2008) Ch.2.3

**Reliability-Failure Case Two:** I receive testimony from a speaker on a topic – say, locations of and routes to get to various streets in East Oxford. I ask her where Marston Street is, to which she knows and gives me the correct answer. I might easily have asked her about another street, about which she would not have known, and would have answered confidently but incorrectly. In this second case we have **a speaker who speaks from knowledge, but accidentally so. She might easily have instead given relevantly-similar false testimony.** There are many variants we can spin on this, for instance Cyril and Maud above.

**Reliability-Failure Case Three.** I solicit information from an honest and knowledgeable speaker, who is extremely reliable about all the topics I might easily ask her about. But I am in an environment where she is the only reliable informant, and there are many wholly unreliable informants around. **It is only by chance that I asked the knowledgeable speaker, and not one of the unreliable ones.**<sup>43</sup>

Variants on this last example abound, not all of them types of testimony. I happen to consult the only accurate travel guide amongst many highly erroneous ones; the only reliable TV news channel among many; the only accurate clock amongst many inaccurate ones... and so forth<sup>44</sup>.

I think Case Three is that in which we are most inclined to ascribe knowledge, despite the failure of **Reliability**, and Case One is that in which we are least inclined to. Why is this? Many writers about testimony have emphasised the idea that by testifying, a knower is able to pass on her knowledge to another person, the recipient. I think we are sure that the recipient's true belief in Case One is not knowledge, because the speaker had no knowledge to pass on. She does not get the goods, because the speaker has no goods to give to her. In contrast, in both Case Two and Case Three, the speaker expresses her knowledge in speaking. I think there is a strong intuition here that the speaker has the goods, she makes available what she has, and so the recipient is able to get them from her – even if she might easily not have been able to. We conceive the situation analogously to the lucky success in getting the needed medicine from the chemist.

This way of describing the situation, and diagnosing our intuitions, is specific to testimony. We could diagnose slightly differently. (Goldman 1979) suggests that the

---

<sup>43</sup> Lackey (2008) Ch.5 contains several examples, each one of which is structurally similar to one of those given here. Her discussion is focused mainly on issues of doxastic responsibility, and she does not explicitly formulate actual-world modal reliability as a matter at issue. Her proposed necessary condition for the gaining of knowledge through testimony, that 'the environment in which B receives A's testimony is suitable for the reception of testimony' is aimed at ensuring modal reliability, being aimed at ruling out Case-Three situations. Thus it is, by the lights of our present proposal, too strong.

<sup>44</sup> Harman's classic example, in which a tourist in a totalitarian-regime country by chance reads the only reliable newspaper, began this genre. See Harman (1973).

processes we regard as yielding justified belief are just the reliable ones. My proposal is that we adapt a version of this idea to give a novel twist to the reliabilist account of knowledge, as follows. We have a folk ontology of reliable sources of belief, thought of by the folk as ways in which we can and often do get knowledge. These include perception, memory and the testimony of other people. These, I now conjecture, are in our folk-ontology of ways you can get knowledge due to their perceived reliability across a broad range of the environments we are used to, and with respect to which we calibrate them. They are, as it were, on our folk-list of approved methods of getting knowledge. (In this respect they contrast with various folk-ways of not getting knowledge – wishful thinking, consulting tea-leaves, and so forth.) Learning that P through being told by someone who knows is a way of getting knowledge that P; and this remains true even if this occurred only due in part to good luck. If this diagnosis and its implications is accepted, then **Reliability** is rejected as a necessary condition for knowledge. But the spirit of reliabilism is maintained, since it is the believed general reliability of a source of belief that gets it onto our approved folk-list of ways in which knowledge is obtained, in the first place<sup>45</sup>. What is dropped is the idea that a method must be reliable in the actual world, the actual situation of a particular believer's belief-formation, in order to yield her knowledge in that situation. We use reliability considerations to fix our list of approved ways of getting knowledge; and then we judge that knowledge is gained in any instance in which one of these folk-approved ways functions normally to yield a true belief – even if a freakily bad epistemic environment means that actual-world modal reliability is not satisfied. Clocks that keep accurate time yield knowledge of what time it is when consulted, and one gets knowledge from consulting a reliable TV news channel or book, or a knowledgeable speaker – even in a situation where one might easily have instead consulted an unreliable one. - I get the goods, even though I might easily have failed to.

Adopting this novel version of reliabilism – **Approved-List Reliabilism** - would not mean accepting as correct the fundamentalist principles K-SUFF and DEFJUS. The internalist requirement of doxastic responsibility still applies, and the debate between fundamentalists versus reductionists over whether this requires positive reasons for acceptance of testimony, or only apt sensitivity to defeaters, remains exactly where it previously stood.

It does have revisionary implications for our view of what it takes to get perceptual knowledge. If all that is needed for getting knowledge is the normal operation of an approved method for getting knowledge, then contrary to what may be described as the received view, dating from Goldman (1976), even when passing through fake-barn country, a normal

---

<sup>45</sup>Goldman (1992) makes a very similar proposal about epistemic justification. However, it seems he still holds to the straight modal reliability condition as necessary for knowledge, that is here questioned. (Goldman, verbal comment, Bled, Slovenia June 2013).

perceiver Roberta comes to know that there is a barn before her on the basis of perception, so long as she has a perceptual capacity for barns in normal, non-freakish situations. If the presence of many fake barns is a one-off, then this does not impugn Roberta's knowledge. Clearly, this suggestion needs much fuller investigation, but I will deviate from my main topic no further in the present essay.

(It matters that the fake-barn set-up is a freakish one-off. If it is a stable feature of Roberta's ongoing environment, the place where she has grown up and that she resides in, that there are both real and fake barns she may easily encounter, and she cannot tell them apart on the basis of perception, then she does not have a perceptually-based recognitional capacity for barns at all. It is plausible that the representational content of perceptual experience is externalistically individuated, so that the concept 'barn' is liable to feature in Roberta's perceptual contents just if she has a perceptually-based discriminatory capacity for barns – that is to say, she can tell them apart from the alternatives she encounters in her usual environment. It is the same with our example of robins/bobbins. I can tell robins apart from the alternatives that show up in my usual environment (my garden), and so I have a perceptually-based recognitional capacity for robins, and the concept 'robin' features in the representational content of my perception, when I spot a robin. When I go on holiday I am still liable to have the perceptual content 'robin' – there is some inertia here - but it is prompted also by bobbins and so I am fooled. If I stay long enough in the new environment, and fail to learn to tell robins from bobbins, my perceptions will no longer bear their former representational content, and will cease to feature 'robin'. It will no longer be true, as it once was, to say of me when I am visually confronted with a robin: it visually seems to her as if there is a robin before her.)

Hesitations over whether Cases One to Three and other intuition-testing scenarios are instances of knowledge do not reflect uncertainty as to the underlying normative facts. One who has a robustly-reliable T-method is better equipped epistemically than one who has a solely actual-world reliable one; one who has a belief that is modally reliable is better epistemically placed than one who forms a true belief in circumstances where she might easily have formed a similar but false one. Our hesitations are over precisely where, in the graded scale of epistemic well-offness, to place the cut-off for knowledge. A richer, less knowledge-obsessed epistemology will not agonise over the placement of the cut-off, but will develop accounts of a range of weaker and stronger epistemic virtues. Attributions of knowledge or its lack are a coarse-grained all-or-nothing snapshot from a more complex normative landscape. They play a central role in our everyday projects and purposes – as regards others, whom to consult, and trust, about a topic (see (Craig 1990)); as regards ourselves, when to act decisively on putative information, and when instead to seek further evidence or second opinion (see (Hawthorne 2004, Stanley 2005); and so forth. Hence

knowledge is a central category in our folk-epistemic thinking. But our pragmatically-oriented attributions of knowledge are driven by various potentially conflicting intuitions, from which different ones will be more salient, and drive our verdict, in different scenarios. Hence our verdicts are interest-driven and highly context-sensitive. We should not expect principled determinacy and stability to characterise our folk-intuitions about the distribution of knowledge.

Thus, for instance, we hesitate as to whether Cyril knows Maud has been working late – he has learned the truth from a folk-approved knowledge source; but his belief is not modally reliable. In fake barn country, a normal perceiver with a recognitional capacity for barns in her usual, more epistemically favourable environments, sees a barn before her and forms belief that she is confronted with a barn; but her belief is not modally reliable, due to a one-off bad-epistemic-luck environment. She enjoys a normal instance of a folk-approved method of getting knowledge; but in abnormal and peculiarly unfavourable circumstances. My aim in this discussion is not to pronounce categorically on whether actual-world modal reliability is necessary for knowledge, but to clarify what is at stake. I am inclined to share the verdict of (Gendler and Hawthorne 2005) that we should not expect to find a unified theory that yields a correct answer covering all cases, or at least no unique all-purpose answer from the folk, to this question. I do maintain that when we turn to testimony, Case Two and Case Three above look like cases of knowledge, so that the standard modal reliability condition **Reliability** appears over-strong.

When you know something, and tell it to me to share your knowledge with me, and I with doxastic responsibility accept what you tell me, you pass on your knowledge to me. This is so even if you or some other nearby person might easily have instead lied to me, and I would still have believed you or her. The account of knowledge I have proposed has this upshot. **All that is needed for getting knowledge is the normal operation of an approved method for getting knowledge.** More strictly, all that is needed for getting knowledge according to the folk, is the normal operation of a folk-approved method for getting knowledge<sup>46</sup>. I have further suggested that the folk-approved-list is selected on the basis of perceived general reliability.

The suggestion made here is only a pointer towards a possible fully-worked out account of knowledge that incorporates reliability in a novel way<sup>47</sup>. Much must be said about what

---

<sup>46</sup> It may be commented at this point that I have made a suggestion about the basis and shape of our everyday attributions of knowledge, what our folk concept entails, rather than about the features that are in fact crucial for knowledge. So be it. Delicate and key methodological issues arise here, which I cannot now address. Goldman (1992) makes important use of a distinction between folk-attributions of knowledge, and correct attributions.

<sup>47</sup> Given what is said above about the unstable context-sensitivity of intuitions about knowledge, the status of such a theory would be explicative, rather than a strict analysis of our folk concept. See

constitutes the ‘normal operation’ of an approved method. There is the generic method, e.g. visual perception; forming belief through acceptance of testimony; addition; and then there is the particular believer’s instantiation of that method. A ‘normal operation’ of the method will involve a competency condition on the believer, and this latter will surely somehow involve reliability of process. The force and point of the present proposal is that this condition will not entail modal reliability – not even method-relativised modal reliability – of the employed method in the actual situation of belief formation, when that is anomalously unfavourable.

My focus has been on whether actual-world method-relativised modal reliability as captured in **Reliability** is necessary for knowledge, and I have claimed the verdict of intuition is that it is not, in the case of testimony. In Cases One to Three the subject fails **Reliability**, but ex hypothesi satisfies a correct doxastic responsibility condition on formation of belief through acceptance of testimony. Another topic, not addressed here, is what a correct such condition amounts to – the familiar issue of Reductionism versus Fundamentalism about testimony. Even doxastic responsibility can seem a very strong condition on gaining knowledge from testimony. – Does one never learn from what one is told, even from a truthful and competent speaker, just because one would not respond with full apt sensitivity to defeaters, if they were encountered? I leave this question for another day<sup>48</sup>.

Is there trouble for **Approved-List Reliabilism** (ALR) over the question whether testimony is indeed a sufficiently reliable source of belief? No. ALR does not turn on the fact of the empirical issue, since what it offers is a story about how our folk-list is derived; so it is folk belief that matters, not the reality of the situation. And the folk count testimony as a way of getting knowledge – it is on the folk-list. In fact, I have argued (Fricker 1994) that the blanket generalisation ‘testimony is generally reliable’ is not a good candidate for a law-like generalisation. It is much more specific sub-types of testimony, typed by speaker, and topic, and – as we have seen in this paper – social context, for which significant generalisations may be sought.

I finish this section with some further thoughts about individuation of methods. We remarked that this is crucial for reliabilism. A T-recipient has an effective filter if she would detect and block relevantly similar false testimony that she might easily have received via the same method. Is the method: receiving any testimony, from any speaker? Or receiving any testimony from that very speaker? or receiving any testimony on similar topics from that/any speaker? I have here assumed that the method is carved broadly, as ‘receiving testimony’; but

---

Kornblith (2002) for an interesting novel take on the nature of a proper philosophical account of knowledge, and the role intuition plays in developing it

<sup>48</sup> Goldberg(2010) supports the idea of some such condition on justified acceptance of testimony. Lackey maintains that ‘in order to acquire knowledge from...testimony...the hearer...has to be a properly functioning or reliable recipient of testimony, where...being appropriately sensitive to defeaters [is] [a] central component of this...requirement.’ (Lackey 2008 p.163)

this might be debated. Equally, what is ‘relevantly similar’ other testimony - suppose my garage mechanic is reliable about cars, but not about driving directions. I would believe him on either. Does this mean my belief about my car’s state derived from trusting his word is not modally reliable? The finer-grained the method, and the narrower the class of relevantly-similar beliefs, the easier it will be, in most cases, for a recipient’s belief to come out modally reliable.

A disjunctivist individuation of methods would render **Reliability** trivially fulfilled in all cases where a normal operation of a method yields true belief.<sup>49</sup> A disjunctivist account of perception holds that in the good case, veridical perception, the subject enjoys a method of finding out about the world, a form of epistemic contact with it, which she simply does not have access to in the bad case, misperception, despite its indiscriminability by the subject from the good case (see (Haddock 2008)). Disjunctivist-style individuation of methods renders reliability trivially fulfilled in any good case. We saw above how one gets knowledge only by chance when one gets the opportunity to use a certain reliable method only by chance. In our example above, I chance to get a glimpse of the burglar, and employ my modally reliable method of visual recognition to recognise him, as he leaves. My belief satisfies method-relativised modal reliability - the bad nearby case in which I get a false belief is one in which I don’t get the chance to employ the method. In the classic fake barn case of Goldman (1976), individuating the method as ‘visual perception’ entails that the same method is involved both in recognising barns, and mis-recognising fake barns as barns; and, in Goldman’s scenario, it is not modally reliable. But if we were to individuate the method as ‘gaining perceptual knowledge’, or ‘perceptually recognising an item’, then it would be trivially reliable. (Compare the case where the bad relevant alternatives are all percept-indistinguishable barn hallucinations.) I am not inclined to disjunctivism, and this would strike me as a cheap trick for reliabilism. However, it may be that our intuition about getting the goods, though one might easily have failed to do so, has a kind of disjunctivist feel behind it. In the case of testimony, **learning that P from someone who knows** is the disjunctivist conception of the epistemic link. That way of getting beliefs is necessarily modally reliable!

## 5. Conclusion: The Need for Social Epistemology

We saw in sect.2 that the empirical studies cited by Shieber do not provide any strong evidence that ordinary T-recipients in their everyday situations fail to embody a T-method with a filter effective for their actual environment. We nonetheless suggested in sect.4 that modal reliability as usually understood is an implausibly strong condition on testimonial

---

<sup>49</sup> One could acquire a true belief via a deviant causal chain, such as those discussed in the 1970s literature on the causal theory of perception, while not instantiating a normal method. See Peacocke (1979).

knowledge. Testimony is a way of getting knowledge, and a person who with doxastic responsibility accepts knowledgeable testimony gets knowledge, even if she narrowly missed a bad case in which she would encounter and be duped by false testimony. We proposed **Approved-List Reliabilism** as a new general theory of knowledge which incorporates reliabilist intuitions, but at a different point.

We observed also that the more powerful a filter a subject's T-method involves, the greater the range of testimonial environments, including very unfavourable ones – ones including accomplished liars and incompetent professionals – in which she will still effectively filter out false testimony, and reliably acquire only true beliefs from testimony. Blind Trust is a reliable T-method in an environment in which one is surrounded only by competent, sincere and epistemically cautious testifiers. Testimony is a two-stage process, and whether its overall upshot is to equip the recipient consistently with only true beliefs depends on the character and more precisely, the aptness of fit between its two stages. We start with an analogy.

Memory is a two-stage process. It involves first the laying down of stored information in memory; second, the retrieval as needed of this information to furnish current belief. So far as I am aware it is not a standard position in the epistemology of memory, that a remembering subject needs to subject her memories to a filter, accepting them at the moment of recall only if they pass some positive epistemic test. The analogy to epistemological Reductionism about testimony is not a current live position regarding memory. (Clearly, globally applicable standards of doxastic responsibility entail that one should be alerted if an apparent memory conflicts with other things one knows.) But the absence of such a recall-point filter means that memory is a reliable source of belief only so long as that there is reliability at the stage in the process when memories are laid down. If humans were such that quite a lot of what got laid down in memory-storage was incorrect, then overall reliability would require an appropriate filter at the recall stage (We may suppose that quite a lot of imaginings got stored as memories, or that dreams regularly got stored as memories of real events.)

What is required, for the whole two-stage memory unit to reliably produce only true beliefs, is that the two stages fit together so that only true beliefs result from the acceptance of apparent-memories. This is ensured by the pairing of an entirely uncritical acceptance of apparent-memories at the moment of recall, with a 100% reliable laying-down process; or by a degree and kind of unreliability at the point of laying-down, so long as this is matched by an aptly tuned filter to detect the duds, at the point of recall<sup>50</sup>.

---

<sup>50</sup> There would also be a failure of reliability, of course, if rather than being accurately preserved once laid down, memories were corrupted or interfered with. I understand empirical work on memory reveals this kind of process. This is not my present topic, and I stick with my toy two-stage model.

The same analysis applies to testimony. Testimony has as first stage the giving of testimony by a testifier, and as second stage its uptake by a hearer. What is needed for overall reliability, is that the filter at the second stage – the filter embodied in the recipient's T-Method – is able to detect the duds, the false testimony, produced at the first stage<sup>51</sup>. Testimony will be a reliable source of beliefs for a recipient in a T-environment E just if her T-filter (narrow T-method) is in this sense well-matched to the testimony she is liable to encounter in E. (We may add that her T-filter should also accept a reasonable percentage of the knowledgeable testimony she encounters. Rejecting everything one is told is a reliable, but not an epistemically advantageous attitude to testimony. See Goldman (1999) ch.4)

The worse the T-environment, the more powerful the recipient's T-filter needs to be, to ensure she accepts only true testimony. The better the T-environment, the weaker a T-filter the recipient can get away with, *salva* reliability. We can marry this observation with what I believe to be the intended overall drift of Shieber's discussion of testimony<sup>52</sup>. If recipients tend to perform poorly at detecting lying and incompetence, then there are two different and potentially complementary ways in which one may seek to improve the overall epistemic situation. One could try to educate and train individual recipients of testimony to do better. Or one could seek to improve their testimonial environment – to find ways to prevent or inhibit the production of the types of false testimony that they sometimes encounter, and are unable to detect. One can work on the first stage, production; and/or on the second stage, reception, to improve the overall reliability or veritistic value of testimony as an epistemic source (see Goldman 1999 ch.4).

The extent to which testifiers in a given institutional or social setting will be careful to obey governing social norms, most notably the norm that one should assert only what one knows to be true, will turn on how effective sanctions against false testimony are in that setting. Disapproval or disappointment of friends and associates gives an informal sanction. In professional settings, institutional norms and procedures can do more to put in place sanctions with serious bite. For instance, the balance between the temptation to massage data to demonstrate a result in the sciences, and the professional sanctions if this were to be found out, is a critical one. The crime of perjury – giving false witness under oath in court – takes the sanction to an extreme, in a situation when the need for truthful testimony is paramount. This thought ends my investigation of reliability and reliabilism in relation to testimony pointing in an apt direction – towards the need and scope for social epistemology, a fast-

---

<sup>51</sup> Just as I abstract from the possibility of corruption of laid-down memories over time, so I here abstract from the scope for the link of testimony to fail due to misunderstanding by the recipient of what the speaker states. Testimony as a link has perfect conditional reliability – it yields only true outputs from true inputs, indeed the very same one, so long as there is no failure of understanding.

<sup>52</sup> As conveyed in personal conversation with Shieber, in symposium at the APA in Chicago, February 2010.

growing and hugely important field pioneered by that veteran philosophical explorer of new territory, Alvin Goldman<sup>53</sup>.

Elizabeth Fricker

Oxford, August 2013

---

<sup>53</sup>Earlier versions of parts of this paper were presented at a conference on Testimony in the Philosophy Department at Nottingham University in May 2012; in the Philosophy Department at Bristol University in October 2012; and at a Social Epistemology conference in Bled, Slovenia in May 2013; as well as at graduate seminars in Oxford. My thanks to audiences for the many comments which helped me to improve the argument. Discussions with John Hawthorne were, as always, an invaluable and equally agreeable help.

Adler, J. E. (1994). "Testimony, Trust, Knowing." *The Journal of Philosophy* **91**(5): 264-275.

Adler, J. E. (2002). *Belief's Own Ethics*. Cambridge, Mass.  
London, MIT Press.

Audi, R. (1997). "The Place of Testimony in the Fabric of Knowledge and Justification." *American Philosophical Quarterly* **34**(4): 405-422.

Bickman, L. (1974). "The Social Power of a Uniform." *Journal of Applied Social Psychology* **4**(1): 47-61.

Bond, C. F., Jr. and B. M. DePaulo (2006). "Accuracy of Deception Judgements." *Personality and Social Psychology Review* **10**(3): 214-234.

BonJour, L. (1980). "Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* **5**(1): 53-73.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, Mass., Harvard University Press.

Burge, T. (1993). "Content Preservation." *Philosophical Review* **102**(4): 457-488.

Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Oxford, Clarendon Press.

Craig, E. (1990). *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford, Clarendon Press.

Dawes, R. M. (1994). *House of Cards: Psychology and Psychotherapy Based on Myth*. New York, The Free Press.

DePaulo, B. M., Stone J.L., Lassiter, G.D. (1985). Deceiving and Detecting Deceit. *The Self and Social Life*. B. R. Schlenker. New York, McGraw-Hill: 323-370.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Oxford, Blackwell.

Ekman, P. and M. O'Sullivan (1991). "Who Can Catch a Liar?" *American Psychologist* **46**(9): 913-920.

Ekman, P., et al. (1999). "A Few Can Catch a Liar." *Psychological Science* **10**(3): 263-266.

Fricker, E. (1987). "The Epistemology of Testimony." *Proceedings of the Aristotelian Society Supplement* **61**: 57-83.

Fricker, E. (1994). *Against Gullibility. Knowing From Words: Western and Indian Philosophical Analysis of Understanding and Testimony.* B. K. Matilal and A. Chakrabarti. Dordrecht

London, Kluwer: 125-161.

Fricker, E. (2006). "Second-Hand Knowledge." *Philosophy and Phenomenological Research* **73**(3): 592-681.

Fricker, E. (2012). "Stating and Insinuating." *Aristotelian Society Supplementary Volume* **86**(1): 61-94.

Gendler, T. and J. Hawthorne (2005). "The Real Guide to Fake Barns: A Catalogue of Gifts for your Epistemic Enemies." *Philosophical Studies* **124**(3): 331-352.

Goldberg, S. (2008). "Testimonial Knowledge in Early Childhood, Revisited." *Philosophy and Phenomenological Research* **76**(1): 1-36.

Goldberg, S. (2010). *Relying on Others: An Essay in Epistemology.* Oxford, Oxford University Press.

Goldman, A. (1976). "Discrimination and Perceptual Knowledge." *Journal of Philosophy* **73**: 771-791.

Goldman, A. (1979). *What is Justified Belief? Justification and Knowledge: New Studies in Epistemology.* G. S. Pappas. Dordrecht

Boston, Mass., D. Reidel Publishing Company.

Goldman, A. (1986). *Epistemology and Cognition.* Cambridge, Mass.

London, Harvard University Press.

Goldman, A. (1992). *Epistemic Folkways and Scientific Epistemology. Liaisons: Philosophy Meets the Cognitive and Social Sciences.* Cambridge, Mass.

London, MIT Press.

Goldman, A. (1999). *Knowledge in a Social World.* Oxford, Clarendon Press.

Graham, P. (2006). *Liberal Fundamentalism and its Rivals. The Epistemology of Testimony.* J. Lackey and E. Sosa. Oxford, Clarendon Press: 93-115.

Greco, J. (2000). *Putting Skeptics in their Place: The Nature of Skeptical Arguments and their Role in Philosophical Inquiry*. Cambridge, Cambridge University Press.

Haddock, A. a. M., Fiona (eds) (2008). *Disjunctivism: Perception, Action, Knowledge*. Oxford, Oxford University Press.

Harman, G. (1974). *Thought*. Princeton, NJ, Princeton University Press.

Hawthorne, J. (2004). *Knowledge and Lotteries*. Oxford, Clarendon Press.

Kornblith, H. (2002). *Knowledge and its Place in Nature*. Oxford, Oxford University Press.

Kraut, R. (1980). "Humans as Lie Detectors: Some Second Thoughts." *Journal of Communication* 30: 209-216.

Lackey, J. (2008). *Learning from Words: Testimony as a Source of Knowledge*. Oxford, Oxford University Press.

Lyons, J. (1997). "Testimony, Induction, and Folk Psychology." *Australasian Journal of Philosophy* 75: 163-178.

McDowell, J. (1994). *Knowledge by Hearsay. Meaning, Knowledge, and Reality*. Cambridge, Mass.

London, Harvard University Press.

Michaelian, K. (2009). "In Defence of Gullibility: the epistemology of testimony and the psychology of deception detection." *Synthese*.

Nozick, R. (1981). *Philosophical Explanations*. Oxford, Clarendon Press.

Owens, D. (2000). *Reason without Freedom: The Problem of Epistemic Normativity*. London, Routledge.

Peacocke, C. (1979). *Holistic Explanation*. Oxford, Clarendon Press.

Ross, A. (1986). "Why Do We Believe What We Are Told?" *Ratio* 28: 69-88.

Shieber, J. (2011). "Against Credibility." *Australasian Journal of Philosophy* 90(1): 1-18.

Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford, Oxford University Press.

Stanley, J. (2005). Knowledge and Practical Interests. Oxford, Oxford University Press.

Stiff, J., Miller, C. (1989). "Explanations for Visual Cue Primacy in Judgements of Honesty and Deceit." Journal of Personality and Social Psychology **56**(4): 555-564.

Welbourne, M. (1986). The Community of Knowledge. Aberdeen, Aberdeen University Press.

Williamson, T. (2000). Knowledge and its Limits. Oxford, Oxford University Press.

Zagzebski, L. (1996). Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge. Cambridge, Cambridge University Press.

Zuckerman, M., et al. (1981). "Verbal and Nonverbal Communication of Deception." Advances in Experimental Social Psychology **14**: 1-59.

Zuckerman, M., et al. (1984). "Learning to Detect Deception." Journal of Personality and Social Psychology **46**(3): 519-528.