Synthese (2025) 206:144
https://doi.org/10.1007/s11229-025-05186-1

ORIGINAL RESEARCH

Testimonial monitoring and the redundancy challenge

Jack L. Herbert[1]

Received: 1 February 2025 / Accepted: 25 July 2025

Abstract
It is widely accepted that the audience has *some* role to play in acquiring testimonial knowledge, but the extent of the epistemic work she must undertake is contentious. At the very least, an audience should not accept testimony that p when she has good reason to believe that $\sim p$. Some take a stronger stance, however, arguing that the audience must demonstrate a counterfactual sensitivity to signs of untrustworthiness. Framed otherwise, some argue that the audience must *monitor* speakers and their reports if they are to acquire testimonial knowledge. In this paper, I present a redundancy challenge to this stronger view. The conclusion I forward is that, if one is to support the view that the attainment of testimonial knowledge requires monitoring, one's defence of this position must appeal to something beyond reliability. As I will argue, if reliability is what matters, then monitoring becomes redundant. For the reliabilist monitoring theorist, this presents a dichotomy: they can prioritise monitoring views over reliabilism, looking beyond reliability to defend monitoring as necessary for testimonial knowledge; or they can prioritise reliabilism over monitoring views, conceding that monitoring is not needed to establish reliable testimonial belief formation. Either way requires a major revision to one's epistemology of testimony.

Keywords Epistemology · Testimony · Monitoring · Reliabilism · Virtue-reliabilism · Social norms · Deception detection

✉ Jack L. Herbert
jackherbertphilosophy@gmail.com

[1] University of Sheffield, Sheffield, UK

Published online: 29 August 2025

1 Introduction

It is widely accepted that the audience has *some* role to play in acquiring testimonial knowledge, but the extent of the epistemic work she must undertake is contentious. At the very least, an audience should not accept testimony that p when she has good reason to believe that $\sim p$. Some take a stronger stance, however, arguing that the audience must demonstrate a counterfactual sensitivity to "signs of untrustworthiness" (Fricker, 1994: 150), or a counterfactual sensitivity to the presence of defeaters (Goldberg, 2007: 168). In this paper, I present a redundancy challenge to this stronger view. The conclusion I forward is that, if one is to support the view that the attainment of testimonial knowledge requires the audience to be engaged in a process of "monitoring", one's defence of this position must appeal to something other than reliability. As I will argue, if reliability is what matters, then monitoring becomes redundant.

In § 2, I introduce monitoring views, contrasting them with anti-monitoring views. The difference between the two, I propose, consists in the monitoring theorist's claim that a failure to monitor thereby blocks testimonial knowledge attainment (Fricker, 1994), (2017), (Goldberg, 2010), (Graham, 2010), (Sperber et al., 2010). Anti-monitoring theorists reject this (Burge, 1993, 1997, 1999), (Simion, 2021). In § 3, I argue that there is a high base rate of speaker honesty in our environment. I forward that our social norms exert a force on speakers which makes the offering of veridical testimony far more likely than the offering of false testimony (Faulkner, 2010), (Graham, 2015), (Simion, 2021). I conclude that, given the high base rate of speaker honesty, we have good reason to believe that testimonial acceptance *without* monitoring will tend towards true belief. Since, for the reliabilist, the epistemic merit of a process consists in its tendency to output true beliefs, those who defend monitoring on reliabilist grounds must concede that *acceptance-without-monitoring* is a reliable process. But if that is so, then monitoring is redundant.

In § 4, I introduce empirical work from deception detection studies to argue that, even if monitoring is reliable, the conditions under which it is reliable are conditions under which a policy of blind trust would also be reliable. In § 5, I look at potential responses the reliabilist monitoring theorist might offer. In doing so, I introduce *reliability consequentialism*: the view that a belief is justified when outputted by a process that has a greater tendency towards true belief relative to alternative available processes. I then show that, while reliability consequentialism rescues monitoring from redundancy, it has unacceptable implications. As such, we must reject it. I conclude by arguing that the reliabilist monitoring theorist faces a choice: they can prioritise reliabilism if they abandon monitoring views, or they can prioritise monitoring views, but must then look beyond reliabilism.

2 Monitoring views and anti-monitoring views

What must we do to acquire testimonial knowledge from a speaker's testimony? Monitoring theorists argue that we must *monitor* speakers, and their reports, for signs of untrustworthiness, or untruthfulness. For the monitoring theorist, failing to monitor therefore blocks the possibility of knowledge. Anti-monitoring theorists reject

this. For them, the attainment of testimonial knowledge is entirely compatible with an absence of monitoring. In this section, I will clarify both views and make clear how I understand monitoring.

First, it is important that we distinguish the descriptive from the normative. Whilst we can engage in philosophical inquiry in both areas, my focus is with the normative. I am interested in the question of whether knowledge necessitates a process of monitoring. Notice that this is different to the question of whether, on the route to knowledge, one of the cognitive steps we take includes a form of monitoring. It may be that the best explanation of our cognitive architecture includes some kind of monitoring. But it is not inconsistent to hold both that we tend to monitor and that testimonial knowledge does not require monitoring. Opening a particularly tight jar does not *require* one to make an odd facial expression, though the making of odd facial expressions often accompanies attempts to open particularly tight jars. To frame things in another way, we might see the debate between monitoring theorists and anti-monitoring theorists as one concerning the *epistemic significance* of monitoring: monitoring theorists take monitoring to be epistemically significant, whilst anti-monitoring theorists do not. This paper argues that monitoring is not epistemically significant in virtue of it securing the reliability of testimonial belief formation. This is because we can establish reliable testimonial belief formation without monitoring.

Throughout this paper, I make a contrast between *acceptance-with-monitoring* and *acceptance-without-monitoring*. One might take this to imply a specific kind of monitoring: something active and intentional. If not, then how could there ever be acceptance-without-monitoring? I recognise this concern. And since I take monitoring to operate, typically, at the subcognitive level, one might suggest that acceptance-without-monitoring must be incredibly atypical. But to make the argument that monitoring is not required for testimonial knowledge, it will be necessary to discuss scenarios wherein an agent might accept a piece of testimony without any monitoring occurring, nevertheless attaining testimonial knowledge in the process. The reader can decide whether such a scenario functions as an argumentative tool or as a genuine possibility. As I see it, the contrast is necessary to demonstrate that one can attain testimonial knowledge without monitoring, even if monitoring is typical.

Finally, it will be beneficial to present a basic description of monitoring before discussing monitoring views (MVs) and anti-monitoring views (AMVs) in detail. Monitoring, as I understand it, is a process, or set thereof, specifically geared towards diminishing the risk of accepting falsehoods. The process of monitoring will involve a sensitivity, therefore, to signs of speaker-insincerity and speaker-incompetence. This might include, amongst other things, the speaker's track-record for reporting truths, fidgety behaviour, tone of voice, and the content of the report. With this description complete, let us begin by looking at AMVs.

Burge can be viewed as defending an archetypal anti-monitoring epistemology of testimony (1993), (1997), (1999). Prior to looking closely at the anti-monitoring element of his account, however, we must first examine his *Acceptance Principle*:

> (AP)– "*A person is entitled to accept as true something that is presented as true and that is intelligible to him*, *unless there are stronger reasons not to do so*" (1993: 281).

Notice that, rather than speaking explicitly of *justification*, Burge speaks of *entitlement* here. On this point, he notes:

> "Although both have positive force in rationally supporting a propositional attitude or cognitive practice, and in constituting an epistemic right to it, entitlements are epistemic rights or warrants that need not be understood by or even accessible to the subject. We are entitled to rely, other things equal, on perception, memory, deductive and inductive reasoning, and on - I will claim - the word of others" (ibid. 272–273).

So, whilst Burge makes a distinction between justification and entitlement, he maintains that they are alike insofar as they both provide *rational support* to beliefs, thereby conferring a positive epistemic status to them. The important difference is that, while justifications require cognitive access, entitlements do not.

Burge starts from the position that perception, memory, and reason are things on which we are rationally entitled to rely. Consider, for example, that when things visually appear to you thus-and-so, your resulting perceptual belief that things are thus-and-so is a candidate for knowledge, assuming you lack reason to the contrary. Testimony, for Burge, is no different (ibid.). We are rationally entitled to rely on the word of others when it is "intelligible" and "presented as true" because, in presenting content intelligibly and as true, we can presume that the source of that content, viz., the testifier, is rational (ibid. 292).

As we see in his AP, Burge states that the entitlement to rely on the word of others can be overridden. This happens when there are stronger reasons not to accept a piece of testimony (ibid. 281). This point can be interpreted in conflicting ways. One might read Burge as stating that an agent's entitlement to accept a piece of testimony is overridden when, irrespective of the agent's access to such reasons, there are stronger reasons not to accept a piece of testimony that an agent *could* or *should have*. However, one might read Burge as stating that an agent's entitlement to accept a piece of testimony is defeated only when the agent *possesses* stronger reasons not to accept it.

I suggest that Burge can be read as defending the latter position, given his comments from "Interlocution, Perception, and Memory":

> "The [acceptance] principle says that the entitlement holds unless there are stronger reasons (available to the person) that override it. It does not say that the person must know there are no stronger reasons; the individual may lack the concept of a stronger reason, and in any case need not rule out the existence of defeaters in advance. It is enough for the individual's being warranted that there are no defeaters; defeaters of the entitlement must be available to him" (1997/2013: 305).

Here, Burge states explicitly that the entitlement to accept testimony is lost only when stronger reasons, which are available to the agent in question, defeat it. If we understand "reasons available to the person" as reasons which an agent has and which she can access, then Burge can be read as defending an anti-monitoring view. And this is because an agent's entitlement to rely on the testimony of others is not overridden by

a failure to engage in some process of monitoring, but simply by the possession of stronger reasons to the contrary.

Burge's account, understood in the manner above, provides us with one way of realising an AMV. According to this position, testimonial acceptance is entitled *by default*, and so, does not depend on any monitoring process. In this sense, Burge defends a default acceptance view, rather than a *monitoring view*. That is, Burge defends a view which allows *acceptance-without-monitoring* to lead to testimonial knowledge, thereby permitting testimonial knowledge *without the need for a monitoring process to operate*. As we have seen, however, the epistemic entitlement that agents enjoy by default can be overridden. And this occurs when an agent has stronger reason not to accept a speaker's testimony.

To make matters clearer, consider the following scenario: a speaker, *S*, reports some proposition, *p*, to her audience, *A*. *A* accepts *S*'s report that *p*, forming the testimonial belief that *p* as a result. *A* has no reason to reject *S*'s report that *p*. So, on Burge's AMV, *A*'s entitlement to accept *S*'s testimony has *not* been overridden; it is entitled by default. As such, *A*'s belief that *p* is a candidate for knowledge, and assuming *S* knows that *p*, *A* comes to know that *p*. Importantly, no monitoring process is *necessary* for *A* to know that *p*. So, for AMV defenders, *acceptance-without-monitoring* can allow for knowledge.

But some figures have argued that AMVs, such as the account advanced by Burge, permit epistemically unacceptable *gullibility* (Fricker, 1994). In response, they have proposed stricter epistemologies of testimony, seeing the audience as having to undertake more epistemic work, by contrast. Take Fricker, for example, who defends a MV according to which "the hearer must always be monitoring the speaker critically" (1994: 154). Views which permit entitled testimonial acceptance without any such monitoring, she argues, permit *blind* or *uncritical* belief "since the hearer's critical faculties are not required to be engaged" (ibid.).

To better understand Fricker's concern, let us consider Goldberg's example of Polyanna (2007: 157). Polyanna is incredibly naïve. She thinks the very best of people, never entertaining the possibility of speaker-insincerity or speaker-incompetence. Imagine that some speaker, Erika, reports some proposition, *p*, to Polyanna. Polyanna accepts Erika's testimony, forming the belief that *p*. While there is no good reason for Polyanna to doubt the truth of what Erika says, even if Polyanna had such reason, she would have nevertheless accepted Erika's testimony. She takes it that, if someone says that *p*, then *p* must be true. Clearly, this is not a good way to go about forming beliefs, and our intuitions reflect this. It would seem mistaken to ascribe testimonial knowledge to Polyanna here. Were we to do so, we would seem to be permitting an extreme form of gullibility. Goldberg recognises this, stating that "[a]n adequate epistemology of testimony must see Polyanna as unjustified in the way that she acquires beliefs through testimony" (2007: 157). Fricker's contention is that AMVs sanction this gullibility; they are "an epistemic charter for the gullible and undiscriminating" (1994: 126).

In response, Fricker argues that "a hearer should always engage in some assessment of the speaker for trustworthiness" (ibid. 145). She forwards that, to acquire testimonial knowledge, audiences must monitor speakers for signs of untrustworthiness (ibid. 150). This is understood as a matter of demonstrating a *counterfactual sensi-*

tivity to signs of speaker untrustworthiness such that "if there were signs of untrustworthiness, [the audience] would register them, and respond appropriately" (ibid.). This might be a matter of recognising observable behaviour, but it might involve appealing to one's background beliefs (ibid.). Through doing this, the audience takes a "critical stance" towards the speaker, "[assessing] her for trustworthiness" (ibid. 154). By adding this condition, Fricker can explain why Polyanna's testimonial belief falls short of knowledge: Polyanna fails to monitor; she fails to demonstrate a counterfactual sensitivity to signs of untrustworthiness.

But suppose that our ability to monitor was so poor that we typically judged those who were untrustworthy to be trustworthy and vice versa. The consequence of this would see us tend to accept testimony from the untrustworthy and reject testimony from the trustworthy. Were this the case, it would be very difficult to understand how monitoring could possibly allow for justified testimonial belief given its high risk of error. Indeed, were monitoring to function in such a manner, it is not clear how this would present much of a contrast to the case of Polyanna. What epistemic merit could monitoring possibly have were it to tend towards failure? As such, I take it that defenders of MVs must regard monitoring as a process, ability, or policy which tends to get things right (ibid.). That is, monitoring theorists must regard monitoring as a *reliable* process in the manner proposed by Goldman (1979/2012).

In advancing his *process-reliabilism*, Goldman argues that justified belief is belief resulting from a process, or set thereof, which tends to produce true beliefs. He writes:

> "The justificational status of a belief is a function of the reliability of the process or processes that cause it, where (as a first approximation) reliability consists in the tendency of a process to produce beliefs that are true rather than false" (ibid. 37).

Applied to the testimonial domain, Goldman's account suggests that *A*'s testimonial belief that *p* is justified when *A*'s belief that *p* is the product of a reliable testimonial belief-forming process. As such, what makes my testimonial belief, say, *that Henry VIII succeeded Henry VII* justified, is that this belief was produced by a reliable testimonial belief-forming process. That is, a process which tends to produce beliefs that are true rather than false (ibid.).

Monitoring theorists, I have suggested, need monitoring to be a reliable process for it to have epistemic merit. But to the question of why monitoring is *necessary* for testimonial knowledge, monitoring theorists may very well disagree amongst themselves. *Reliabilist* monitoring theorists, I propose, will say something along the following lines: monitoring is necessary for testimonial knowledge because monitoring helps to secure the reliability of the testimonial belief-formation process. Testimonial belief formation *without* monitoring would be unreliable, and so, would fail to produce beliefs which are candidates for knowledge. In short, testimonial belief formation must involve monitoring for it to be reliable. This broad view is the focus of the current paper. I will therefore be operating within a reliabilist framework throughout.

Let us now consider Goldberg's process-reliabilist MV. According to this view, an audience, *A*, must demonstrate a "'counterfactual sensitivity' to the presence of defeaters" in order to know that *p* from a speaker's, *S*'s, testimony that *p* (2007: 168).

That is, *A* must demonstrate a sensitivity such that if there *were* defeaters, she *would* respond to them, irrespective of whether there *are* defeaters in the current scenario. This blocks the possibility of knowledge not only for the audience who has excellent reason to reject *S*'s testimony, but also for the audience who is disposed such that, were they in possession of such reason, they would fail to respond to it.[1] Framed otherwise, this condition states that testimonial knowledge can only be gained via a reliable testimonial belief-formation process *which involves monitoring*. Insofar as Goldberg supports this condition, he supports a monitoring view.

How exactly an audience demonstrates a counterfactual sensitivity to the presence of defeaters is another matter. Goldberg suggests that coherence-monitoring processes might play this role (ibid. 169). The thought here is that there are "subcognitive" coherence-monitoring processes which are constantly running (ibid. 171). When a conflict is detected between the received testimony and the background beliefs of the audience, this might "give rise to articulate reflection" (ibid. 170). When that happens, the incoming testimony might be rejected, but it may also be that the audience revises her background beliefs. At any rate, by performing this constant coherence-monitoring, audiences are engaged in monitoring insofar as they subcognitively check for inconsistencies whenever they receive testimony.

Turning to the work of Sperber et al. (2010) and Graham (2010), we see yet further ways to make sense of Goldberg's counterfactual sensitivity requirement. Sperber et al., for example, suggest that "humans have a suite of cognitive mechanisms for epistemic vigilance, targeted at the risk of being misinformed by others" (2010: 359). Some of the mechanisms involved might, they suggest, draw on "some of the capacities used in selecting partners for cooperation, which include moral evaluation, monitoring of reliability, and vigilance towards cheating" (ibid. 372). And Graham implies that we have a more general "filtering" system, involving a range of filtering processes (2010). These filters involve "different sensitivities", such as coherence-checking, track-record information, manner of presentation, and more (ibid. 152–153). We see then that there are various ways to make sense of the underlying mechanisms for monitoring.

I will now make my position on monitoring clearer. I will understand monitoring to function as part of a testimonial belief-formation system which operates, typically, at the subcognitive level. The system, as I take it, involves a range of "mechanisms", or "filters", specifically aimed at diminishing the risk of accepting falsities. These processes may be domain-general, but there is reason to believe that at least some of these processes will be domain-specific. Importantly, we can make a distinction between "world-directed" monitoring processes, and "self-directed" monitoring processes.

[1] The audience who accepts that *p* when everything they know makes *p* highly improbable fails to know that *p* because, amongst other things, they have a defeater for *p*. The audience who accepts that *p* when they have no reason to reject that *p* but is disposed such that they would neglect such reason fails to know that *p* because, as Goldberg puts it, the audience's acceptance of *p* was not "the outcome of a process that exhibited a 'counterfactual sensitivity' to the presence of defeaters" (ibid.). As such, Goldberg's proposal does not allow Polyanna to gain justified testimonial belief or knowledge, given her manner of forming testimonial beliefs.

World-directed monitoring processes involve the scanning of information in one's environment, *beyond one's own mind*. They involve a sensitivity to considerations such as the tone of the speaker's voice, her body language, and so on. A sensitivity to such considerations requires the operation of world-directed processes insofar as it requires the audience to engage with her environment, rather than simply with her own mental states. Consider, if your fidgety behaviour is to function as a sign of untrustworthiness for me, then I must be sensitive to my immediate environment to gain that reason. Contrast this with the case in which you tell me that p when I have very good reason to believe that $\sim p$. In this instance, I need only demonstrate a sensitivity to my own mental states to gain a reason to treat your testimony as suspect. This distinguishes what I call STRONG MONITORING (SM) from WEAK MONITORING (WM). As I see it, advocates of SM take monitoring to involve world-directed and self-directed processes.[2] Advocates of WM take monitoring to involve *only* self-directed processes.[3]

Given my understanding of monitoring, one cannot fail to monitor a speaker, or her testimony, much as how one might fail to check one's bag for one's keys. It may turn out that monitoring, so conceived, fails to operate only in atypical cases (for example, those involving brain damage). This is in keeping with the reliabilist's understanding of other belief-forming processes– for example, those concerning perception– and does not render the current inquiry purely descriptive. To be clear, I am not interested specifically in the question of which steps we complete cognitively on our way to testimonial knowledge. As noted previously, some of the steps may well be epistemically unimportant. Rather, I am interested in the question of whether monitoring should be considered as a *necessary* step, irrespective of the frequency of our engagement in it, on our way to testimonial knowledge. I will argue that, if it is a necessary step, it cannot be simply because it secures reliability in the testimonial belief-formation process.

3 The norm of truth telling

Numerous figures have recognised that the reliability of a given belief-forming process depends, in some sense, on certain factors within the environment in which it operates (Goldman, 1979/2012: 43), (Plantinga, 1986: 12–14), (1988: 33–34), (Sosa, 1991: 276), (Lyons, 2013: 17–19). The reliability of our visual belief-forming processes, for example, is indexed to environments in which we can see without hindrance, where the lighting is not too dark, and so on. This makes sense since our ancestors would have typically lived in environments like ours, visually. As such, the reliability of human vision is indexed to a certain kind of lighting, amongst other things. So, whilst we might tend towards false belief in very dimly lit environments, this says nothing about the reliability of our visual belief-forming processes per se. Rather, it says something about the reliability of our visual belief-forming processes *in a given environment*.

[2] See Fricker (1994), (2017), Graham (2010), Sperber et al. (2010).

[3] See Goldberg (2007), (2010).

In making claims about the reliability of a given belief-forming process, it is important to look at the environment in which it is operating. If we are ascribing knowledge to an agent, we are implicitly claiming that the environment in which that agent's relevant belief-formation process is operating allows for reliable belief formation. Consider, for example, that if we ascribe to an agent the perceptual knowledge *that there is a red book on the table*, then we must implicitly claim that the agent's environment permits the reliable operation of her visual belief-formation processes. This is not to suggest that the converse is true. One might fail to know for myriad reasons. However, it is important to recognise that if an agent's belief is outputted by a belief-formation process operating in an environment that does not permit reliable belief formation, then it would be mistaken to ascribe to her knowledge. This is why a very foggy environment can block visual beliefs from counting as knowledge: it negatively impacts the reliable operation of visual belief-formation processes.

Since reliability is indexed to the environment, it is important to consider the environment in which the testimonial belief-formation process operates when reflecting on the necessity of monitoring. The reliabilist monitoring theorist, as I have suggested, is committed to the claim that monitoring is necessary to achieve reliable testimonial belief formation. The anti-monitoring theorist rejects this. To support the claim that monitoring is redundant in the reliabilist framework, I will argue that the environment in which this belief-formation process operates is one in which speakers typically offer veridical testimony. Due to this, audiences do not need to engage in the kind of epistemic work suggested by Fricker (1994), (2017), Graham (2010), and Sperber et al. (2010). Subcognitive or not, reliable testimonial belief formation does not require monitoring. This can be explained, in part, by our *social norms*.

Faulkner notes that, in cases involving the behaviour of one agent towards another, the violation of social norms can be seen as having "three attitudinal dimensions" (2010: 130). When an agent, *S*, violates a social norm, wronging another agent, *A*, in the process, that violation can induce in *S* feelings of shame, guilt, or embarrassment; it can cause *A* to feel resentment towards *S*; and it can provoke third parties to disapprove of *S* and her conduct (ibid.). Faulkner argues that, where we see these "hallmark emotions" typically provoked, we have good reason to believe that a social norm not only exists but has been violated (ibid.). And so, if we see an instance of this, we should judge that a related social norm exists.

Consider a case in which a dear friend, *S*, has entrusted you with some very personal information, asking that it be kept private. Another friend of yours, *A*, asks you about *S* at a later time, believing something to be troubling *S*. This is a difficult situation for you. You try to get out of it, but *A* is incredibly persistent. Do you tell *A* what *S* said to you in confidence? You decide, reluctantly, to make something up to get *A* to stop pressing you for details. In doing so, you feel somewhat guilty. If *A* were to learn that you were deceptive, she would surely feel resentment. Moreover, others might take issue with your deceptive conduct, notwithstanding the fact that you were put in an uncomfortable position. If you had told *A* what *S* told you in private, however, then *S* would likely feel resentment, you would be subject to disapproval from others, and you would feel shame for betraying *S*'s confidence. Following Faulkner, I argue that this gives us reason to believe that there is a norm of *promising* and, more importantly for our purposes, a norm of *truth telling* (ibid.).

According to Elster, "social norms are enforced by members of the general community" (1989: 100). We can understand how this is achieved by considering the responses community members give to those who violate the norm in question. As noted, failing to tell the truth can cause one's audience to feel resentment towards them, and can provoke feelings of disapproval from third parties. Being subject to a display of these emotions is unpleasant in itself for the overwhelming majority, but further penalties might be incurred. One might, for example, lose social prestige, be denied assistance from conspecifics, be physically harmed, or be ostracized (Graham, 2015: 252). And so, the display of these emotions from others, and the risk of associated penalties, enforces the norm of truth telling and exerts on speakers a pressure to tell the truth. Indeed, the mere *anticipation* of such responses might help to sustain norm-conformant behaviour. I might anticipate, for example, that *if* I were to tell you that p when p fails to obtain, then you would resent me if you learnt that $\sim p$.

Whilst the risk of punishment obviously motivates one to abide by social norms, norms should not be understood as motivating purely due to such threats. As Elster writes, "[w]hen norms are internalized, they are followed even when violation would be unobserved and not exposed to sanctions" (1989: 104). Failing to tell the truth can lead *me* to have feelings of shame, guilt, and embarrassment. And insofar as I feel this in response to an instance of me violating the norm of truth telling, I help to enforce the norm. This point is important as it suggests that, even when a scenario arises in which there is reason to believe that the violation of a norm might not incur some clear penalty, or might not even be recognised, we can still feel pressure to abide by that norm *due to internalization of it*. As Graham notes, "[w]hen we internalize a norm, we find it intrinsically motivating; our preferences change. We conform because we think it's the right thing to do, because we are supposed to do it" (2015: 253). And this gives us a further way to understand the force speakers feel to tell truths.

By way of sanctions and norm internalization, we help to enforce the norm of truth telling and, in so doing, put pressure on ourselves, and others, to tell the truth. Of course, we know that sometimes people lie, and other times they just make honest mistakes. I do not deny that, in some cases, it might be more beneficial for a speaker to attempt a deception than to be honest. I also recognise that, in some cases, a speaker might not care that they will be subject to scorn. But such examples tell us only that people sometimes violate the norm of truth telling, not that the norm fails to exist, or that it fails to exert pressure on speakers. My claim is not that speakers never violate the norm of truth telling, it is that there exists a norm of truth telling which makes deviation from that norm *atypical*. And insofar as deviations from that norm are atypical, we should judge that most speakers abide by the norm. That is, *most speakers tell the truth*. Framed otherwise, we can say that the base rate of speaker-honesty, at least in our community, is very high.

Now, if there is a very high base rate of speaker-honesty, then we have reason to believe that a given instance of testimonial acceptance will most likely be an instance of accepting a truth. And if that is right, then very little is needed to establish a reliable testimonial belief-formation process. Consider an analogy: if you put your hand into a jar of 100 marbles, 90 of which are red and 10 of which are blue, then a given instance of picking a marble blindfolded will most likely also be an instance of pick-

ing a red marble. Statistically speaking, it is just far more likely that you pick a red marble. Would a process of looking at the marbles, carefully scrutinizing each one prior to picking it be a process that tends to select red marbles? Of course. But is this process *necessary* to achieve that tendency? The answer is no. If what you want is a tendency to select red marbles, then carefully scrutinizing each marble is simply not needed; it is redundant. The same is true in the testimonial case. It is far more likely, if the base rate of speaker-honesty is very high, that you will receive veridical, rather than false testimony. Will a process of monitoring achieve reliability? Yes. But is it *required* to achieve reliability? No. All that is needed is a process for comprehending what is said and then storing it, in addition to the high base rate of speaker-honesty. If what you want is reliability, then monitoring is redundant. The high base rate of speaker-honesty suffices to secure this.

I have argued that violations of the norm of truth telling are atypical due to the force exerted on speakers by norm internalization and the risk of sanctions from other community members. Insofar as violations of the norm are atypical, the base rate of speaker-honesty is very high. To reinforce this, I will briefly discuss the work of Simion (2021), who offers another route to this conclusion.

Simion has argued that speakers rarely deviate from the *knowledge norm* of assertion, which sees epistemically permissible assertions as assertions which are *known* by the speaker (ibid. 909). While it might benefit speakers to refrain from telling the truth whenever it suits them, the existence of the knowledge norm "changes the utility profile of testimonial exchanges" (ibid. 910). This is to say that the existence of the knowledge norm makes norm conformity beneficial for speakers. One can, for example, gain a good reputation, secure an increased chance of conspecific aid, and lower the risk of group disapproval (ibid.). Insofar as the utility profile of testimonial exchanges is affected by the existence of the knowledge norm, it will be rational, by default, for speakers to abide by that norm (ibid.). Moreover, it will be rational, by default, for hearers to believe what speakers say. And if this is correct, and we accept that agents tend to be rational, then we have reason to believe that speakers will tend to tell the truth, and so, that the base rate of speaker-honesty will be very high.

The arguments provided in this section support the claim that there is a social norm governing speaker behaviour. The consequence is that speakers tend to abide by that norm, making violations atypical. As such, I argue, we have good reason to believe that there is a high base rate of speaker-honesty in the environment we inhabit. And if there is a high base rate of speaker-honesty, then a given instance of testimonial acceptance will likely lead to true testimonial belief formation. Monitoring is not needed to establish reliable testimonial belief formation on this picture because, in our environment, speakers tend to abide by social norms governing assertion. And so, if monitoring is not required for reliable testimonial belief formation, it is not clear why it would be required for testimonial knowledge *on reliabilist grounds*. If we see the epistemic value of a process as exhausted by its tendency to produce true beliefs, then monitoring is redundant, given the picture outlined in this section.

4 Deception detection

I will now reinforce the claims of § 3 by introducing the empirical findings of the *deception detection* studies. Before doing so, however, I will clarify my goal here. The deception detection studies are psychological experiments which test our ability to detect deceptions in real time. Simion has used such findings to argue against Graham's view, according to which the attainment of testimonial knowledge requires the operation of a reliable "filtering" process (ibid. 897–898). I will not adopt a stance on that particular issue in this paper.[4] My aim is not to argue that monitoring is *unreliable*, rather it is to show that, even if we allow that monitoring is reliable — despite the empirical findings — then it is nevertheless redundant because the conditions under which monitoring is reliable are conditions under which a policy of blind trust is reliable too. Let us now proceed.

According to Vrij, the typical format of a deception detection study runs like so:

> "[O]bservers (normally undergraduate students) are given short video fragments of people they do not know who are either telling the truth or lying. They are asked to indicate after each fragment whether the person (often called the sender) was telling the truth or lying. Typically, half of the senders are truth tellers, and half are liars" (2008: 147).

If we have a reliable monitoring system, then we should expect to see accuracy rates which reflect this.

Vrij writes that, "[i]n such a study, simply guessing whether the sender spoke the truth or lied would result in correctly classifying 50% of the truths (truth accuracy rate) and 50% of the lies (lie accuracy rate), resulting in a total accuracy rate of 50%" (ibid.). Additionally, classifying each report as "truthful" would yield a truth accuracy rate of 100%, a lie accuracy rate of 0% and a total accuracy rate of 50%. Conversely, classifying each report as "deceitful" would yield a truth accuracy rate of 0%, a lie accuracy rate of 100% and a total accuracy rate of 50% (ibid.). We should expect to see a total accuracy rate considerably greater than 50% if monitoring is reliable.

The findings from the deception detection studies present a challenge of sorts for monitoring theorists. In their meta-analysis, Bond and DePaulo write that:

> "Despite decades of research effort to maximize the accuracy of deception judgments, detection rates rarely budge. Professionals' judgments, interactants' judgments, judgments of high-stakes lies, judgments of unsanctioned lies, judgments made by long-term acquaintances– all reveal detection rates within a few points of 50%. We wonder if it is premature to abort the quest for 90% lie detection and accept the conclusion implied by the first 384 research samples– that to people who must judge deception in real time with no special aids, many lies are undetectable" (2006: 231).

[4] It is worth stating nevertheless that, insofar as the empirical data give us reason to doubt the reliability of some world-directed monitoring processes, the SM theorist should offer a response to the concerns introduced by the experimental findings.

In short, their conclusion suggests that monitoring is not reliable.

Before proceeding, I want to pause and recognise the reservations that one may have with respect to the empirical findings. If they do test any monitoring processes, one might suggest, surely their scope is very narrow. It is not clear, for example, how the deception detection studies could test a process of coherence-checking. And so, if one wishes to use the deception detection data to argue against the reliability of monitoring, one will need to direct that challenge to the SM theorist specifically. I think this is correct.

It is, however, hard to dispute that the deception detection studies test our ability to detect deception on the basis of *observable behaviour*. The studies examine our ability to discriminate between lies and truths on the basis of fidgety behaviour, eye movement, tone of voice, and things of this nature. What's more, the findings give us reason to believe that any monitoring process we have directed at such considerations is not reliable. If one is sympathetic to SM, one will need to respond to this concern, arguing why it fails to show that world-directed monitoring processes are unreliable.

Whilst I think that the empirical literature gives us reason to doubt the reliability of at least *some* monitoring processes, I am not interested in defending the claim that monitoring is unreliable. Instead, I have introduced the empirical data to argue that *if* monitoring is reliable, then acceptance-without-monitoring must also be reliable. And so, if the epistemic merit of a process is exhausted by its reliability, then monitoring is entirely redundant.

Here, one might ask: if the experimental findings suggest that our deception detection accuracy scores hover around 50%, how could one defend a claim which relies on the possibility of monitoring being reliable? As previously argued, the reliability of a process is indexed to the environment in which it operates. As such, I will explore the possibility that monitoring is reliable, but *outside of the experimental setting*. My aim will be to show that, even if monitoring is reliable, it adds no further epistemic value to the testimonial belief-formation process, given that acceptance-without-monitoring already constitutes a reliable testimonial belief-formation process. With my aim here clarified, let us continue.

In his critique of Fricker's monitoring view, Michaelian exercises charity in taking the typical accuracy rate of participants in the deception detection experiments to be 57% (2010: 415). But he notes, importantly, that focusing on this statistic alone is misleading. The issue is that it fails to recognise two phenomena: (i) truth-bias; (ii) veracity effect. Regarding the former, Levine et al. comment that "independent of actual message veracity, individuals are much more likely to ascribe truth to other's messages than deceit" (1999: 126). That is, we have a bias such that we are more likely to judge reports as truthful than deceitful. The presence of this truth-bias helps to drive, what Levine et al. dub, the "veracity effect": "truths are identified with greater accuracy than lies" (ibid. 139). The thought here is that if I am disposed to judge more reports as truthful rather than deceptive, I am more likely to score higher on truth accuracy than lie accuracy. Consider, if you toss a fair coin 100 times and I guess heads more often than tails, the probability of me guessing heads when the coin lands on heads is greater than the probability of me guessing tails when the coin lands on tails. In such a case, we should expect to see a "heads accuracy score" greater than that of a "tails accuracy score". Moreover, if the coin landed on heads more often in

that sequence of 100 tosses, thereby presenting a sequence with a higher base rate of heads, we should also expect a higher "overall accuracy score" given the increased likelihood of guesses correctly aligning with outcomes. If this is right, we should anticipate that a change in the base rate of speaker-honesty, should bring about a change in overall accuracy of deception detection. Levine et al. tested this hypothesis.

Levine et al. had two students record a series of statements, 6 truthful and 6 deceptive. They then made three sequences from those recordings. In sequence 1, four of six deceptive statements were deleted, yielding a 75% speaker-honesty sequence. In sequence 2, four of six honest statements were deleted, yielding a 25% speaker-honesty sequence. And in sequence 3, two honest and two deceptive statements were deleted, yielding a 50% speaker-honesty sequence. The hypothesis Levine et al. forwarded was that "*truth-lie base rates will significantly affect detection accuracy such that accuracy will be the highest in a 75% honest condition*, *significantly lower in a 50% honest condition*, *and lowest in a 25% honest condition*" (ibid. 137). They called this the base-rate effect. Their findings were consistent with this hypothesis (ibid. 139). They found that an increase in the base rate of speaker-honesty is positively correlated with an increase in overall detection rate. This point is key.

When we increase the base rate of speaker-honesty enough — to around 75% and upwards — overall detection rates are not so bad. The higher the base rate of speaker-honesty, the better we perform. This gives an attractive response to the concern that monitoring is not reliable. The monitoring theorist can say that the base rate of speaker-honesty in the deception detection studies is too low. But when we increase it to a rate more representative of the non-experimental environment, we perform well. The issue is that *the same is true of a blind trust approach*. That is, the same is true of adopting a policy whereby one simply accepts whatever one is told. As Michaelian puts it "if the base rate of honesty is sufficiently high for the monitoring process to be reliable, then it is sufficiently high to render the blind trust process reliable" (2010: 419). And if we hold that the justificatory status of beliefs is determined by the reliability of the processes outputting them, then it is unclear why an AMV is any less attractive a position than a monitoring view. If one wants to resist this move, perhaps speculating that the base rate of speaker-honesty is not *that* high in "everyday situations", then whilst blind trust is no longer a reliable process, neither is monitoring. But if monitoring is only reliable under those conditions where blind trust is also reliable, and what we value epistemically is reliability, then why should monitoring be considered *necessary* for knowledge? My response is that it shouldn't.

5 The most reliable process: reliability consequentialism

The reliabilist monitoring theorist must say why monitoring is not redundant. They must explain why it is *necessary* for knowledge. As we have seen, the suggestion that monitoring is necessary because it is required to secure reliable testimonial belief formation fails. This is because our environment's high base rate of speaker-honesty helps to establish reliable testimonial belief formation even without the operation of monitoring processes. So, what else might the MV advocate say? One possible response goes like this: whilst acceptance-without-monitoring might be reliable,

acceptance-with-monitoring is *more* reliable. Through monitoring, we form testimonial beliefs in the *most reliable manner available to us*. And insofar as monitoring has a greater tendency towards the truth, it is an *epistemically preferable process*.

But how might the monitoring theorist claim that monitoring is *more reliable* than acceptance-without-monitoring given the empirical findings? As I see it, the MV advocate will have to take issue with something very basic to the deception detection experiments. Perhaps the studies fail to represent what typical testimonial exchanges look like. Perhaps they test merely a subset of the processes involved in monitoring. Perhaps most of our testimonial interactions are with people about whom we know at least a little. I will not devote time to investigating possible defences for this claim. Moreover, I recognise that the empirical findings might not affect the WM position at all. For the sake of argument then, I will simply proceed on the understanding that one can defend the claim *that acceptance-with-monitoring is more reliable than acceptance-without-monitoring.*

The reliabilist monitoring theorist might begin their reply by first remarking that belief-forming processes which are reliable are epistemically valuable, while belief-forming processes which are unreliable are not. Forming beliefs about the saltiness of foods on the basis of taste is a reliable process, for example, while forming beliefs about the saltiness of foods on the basis of sound is not. Since, for reliabilists, it is the reliability of a process that gives it epistemic merit, we can view the former process as having epistemic merit and the latter as lacking it. The next move the reliabilist might make is to suggest that, if we accept that the epistemic merit of a process is exhausted by its reliability, we should not only prefer reliable to unreliable belief-forming processes, but we should prefer belief-forming processes which are more reliable than alternatives. So, in the scenario where one can form the belief that *p* by means of process$_1$, which outputs true beliefs 90% of the time, or process$_2$, which outputs true beliefs 70% of the time, we should regard process$_1$ as more epistemically valuable, and thus, *epistemically preferable* to process$_2$. And if we regard one process as epistemically preferable to another, we should use that process over alternatives. Next, we just need the claim that monitoring is preferable to acceptance-without-monitoring, and we reach the conclusion that, since acceptance-with-monitoring is epistemically preferable to acceptance-without-monitoring, we should monitor in all scenarios where it is an available alternative. Given that monitoring, as I have understood it, consists in a range of subcognitive processes, this will include the overwhelming majority, if not all, instances of receiving testimony.

But there is an issue here that requires further attention. We can make a distinction between something being *epistemically preferable* and something being *epistemically necessary*. Consider the following case: Fatima asks Katya what she had for dinner three nights ago. Katya reflects for a moment and then responds, "pasta". As Fatima is fully aware, Katya is on a diet and is recording all her meals in a journal. What is the epistemically preferable way for Fatima to form a belief about what Katya had for dinner three nights ago? She could rely purely on what Katya says, but she could supplement her testimony with the journal entry. Surely, the risk of forming a false belief is greater in the case wherein Fatima relies *only* on Katya's testimony. After all, she might be misremembering. So, the epistemically preferable option would see Fatima use both Katya's testimony *and* her journal entry. But it seems mistaken to

suggest that this is epistemically *necessary*, even if we allow that it is epistemically *preferable*. If Katya knows that she had pasta for dinner three nights ago and says this to Fatima, surely Fatima comes to know that Katya had pasta for dinner three nights ago on the basis of her testimony. But this would suggest that something can be epistemically preferable without being epistemically necessary. And if that is the case, then the monitoring theorist is unable to support the claim that monitoring is necessary for testimonial knowledge because it is more reliable, and thus, epistemically preferable to acceptance-without-monitoring. Even if we allow that acceptance-with-monitoring is epistemically preferable to acceptance-without-monitoring, this does not allow one to move to the claim that acceptance-with-monitoring is epistemically necessary.

The simplest way to get from an epistemically preferable process to an epistemically necessary process is to revise our understanding of knowledge, demanding instead the *most reliable process*. Call this *Reliability Consequentialism*. Reliability Consequentialism (RC) holds that:

> RC: *S*'s belief that *p* is justified at *t* only if *S*'s belief that *p* was produced by a belief-forming process that has the greatest tendency to output true beliefs relative to alternative belief-forming processes available to *S* at *t*.

While RC differs from Goldman's process-reliabilist thesis as presented in § 2, it is clearly a reliabilist position, nevertheless. Indeed, one might draw a parallel between RC and one of Goldman's early reliabilist proposals:

> "If *S*'s belief in *p* at *t* results from a reliable cognitive process, and there is no reliable or conditionally reliable process available to *S* which, had it been used by *S* in addition to the process actually used, would have resulted in *S*'s not believing *p* at *t*, then *S*'s belief in *p* at *t* is justified" (1979/2012: 46).

The key difference between RC and this proposal, however, is that Goldman offers a *satisficing* theory, whereas RC is a *maximizing* theory. That is, Goldman's reliabilism permits justified belief so long as *S*'s belief that *p* is the product of a process which surpasses a given threshold (and, given the suggestion above, there is no other process available to *S* which would have resulted in her believing otherwise). RC, on the other hand, permits justified belief only when *S*'s belief that *p* is the product of *the most reliable process*. Goldman's process-reliabilism therefore allows "suboptimal" yet reliable processes to produce justified beliefs while RC does not; only the best is good enough.

Can a defender of RC push back against the redundancy challenge I have introduced? To do so, the RC advocate would first need to establish that monitoring is indeed more reliable than straightforward acceptance-without-monitoring. While the empirical findings might give us reason to be suspicious of this claim, let us assume as much for the sake of argument. Let us continue on the understanding that, while acceptance-without-monitoring might permit reliable testimonial belief-formation, acceptance-with-monitoring would secure greater reliability. According to the RC thesis, monitoring is therefore *necessary* for knowledge. If one fails to monitor, one

is thereby blocked from attaining testimonial knowledge. So, through defending RC, one can support monitoring as necessary for testimonial knowledge on purely reliabilist grounds. The issue, however, is that we have very good reason to reject RC.

Firstly, RC suffers from the "best of the bad lot" objection. Consider what Dunn and Ahlstrom-Vij dub, the *epistemically tragic world* "where *all* processes are unreliable, and the maximally reliable one is simply slightly more reliable than the others" (2017: 191). According to RC, the maximally reliable, yet still unreliable process permits knowledge attainment. But this would entail that the belief-forming process which tends to output true beliefs 5% of the time permits knowledge because it is more reliable than available alternatives which output truths, let's say, 3% of the time. Clearly, we do not want to allow this possibility. So, let us revise RC:

> RC*: *S*'s belief that *p* is justified at *t* only if *S*'s belief that *p* was produced by a reliable belief-forming process that has the greatest tendency to output true beliefs relative to alternative belief-forming processes available to *S* at *t*.

RC* holds that: (1) to know that *p* requires one's belief that *p* to be outputted by a reliable belief-forming process; (2) to know that *p* requires one's belief that *p* to be outputted by the most reliable available belief-forming process, relative to alternative available processes. As such, RC* circumvents the "best of the bad lot" objection, since knowledge requires reliability, not *merely* the most reliable process. However, RC* also has unacceptable implications.

Let us consider another example. Imagine a not-so-distant future where a new piece of technology is widely available. This new technology, appearing to be an ordinary pair of spectacles, allows users to form visual beliefs with greater accuracy than typical eyesight. In short, this device is more reliable than human vision, by and large. Suppose that this technology is made widely available to the public. That we could easily access such technology clearly does not affect the reliability of our unenhanced visual belief-forming processes. Human vision, typically, is *reliable enough*. Surely, we want to allow that, even in this scenario, human vision typically allows for visual knowledge. But if RC* is correct, then the widespread availability of this technology would disqualify visual beliefs produced by human eyesight alone as candidates for knowledge. RC* commits one to holding that the widespread availability of this technology prohibits visual knowledge via human eyesight, since human eyesight would be less reliable than the new technology (though still reliable). And there are doubtless further examples similar in nature to this. The availability of calculators, for example, could be seen as a present-day case. That so many of us can access calculators on our phones so easily surely does not rule out cases of mental arithmetic as allowing for knowledge, despite mental arithmetic plausibly being less reliable than calculators. But again, this is what RC* entails.

We have good reason to reject RC and RC*. And, if we are committed to reliabilism, this would seem to push us back to where we started, viz., to Goldman's satisficing process-reliabilist position. Without any clear way to respond to the anti-monitoring theorist's claim that monitoring is redundant, the reliabilist monitoring theorist appears to lack the tools to rule out acceptance-without-monitoring as a reliable process.

In search of a response, might one perhaps defend monitoring by way of *virtue-reliabilism*? If we follow the virtue-reliabilists, such as Greco (1993), (2003), Sosa (2007), (2015) and Pritchard (2010), viewing monitoring as a kind of "skill", "ability", or "competence" through which we achieve reliable belief formation, then perhaps we can support MVs without giving up on reliabilism. To better understand the position, let us consider the following passage from Greco:

> "When we say that S knows *p*, we imply that it is not just an accident that S believes the truth with respect to *p*. On the contrary, we mean to say that S gets things right with respect to *p* because S has reasoned in an appropriate way, or perceived things accurately, or remembered things well, etc. We mean to say that getting it right can be put down to S's own abilities, rather than to dumb luck, or blind chance, or something else" (2003: 116).

So, the virtue-reliabilist monitoring theorist will argue that it is through monitoring that one "gets it right" by way of their own abilities.

But this move is too quick. To suppose that anti-monitoring theorists are forced to regard the formation of true testimonial beliefs via acceptance-without-monitoring as a matter of "dumb luck" or "blind chance" is uncharitable. Anti-monitoring theorists, I argue, have the tools to make sense of testimonial knowledge attainment in virtue-reliabilist terms if they appeal to some competence other than monitoring that accounts for an audience's "getting things right". The monitoring theorist will, of course, regard monitoring as the epistemically significant ability accounting for one's getting things right, but there is no reason to see this as the *only* available option.

I suggest that, if the anti-monitoring theorist views *comprehension* as a kind of competence, they can maintain that monitoring is not necessary for knowledge. And why should we not see comprehension as a competence? Indeed, this is Burge's view (1993), (1999). According to Burge, our capacity to comprehend content is what "preserves" and "enables" the content of utterances from speaker to hearer (1993: 293). Without such a capacity in place, communication would fail.[5] For it is through comprehending a speaker's veridical testimony and forming the corresponding belief that an audience succeeds in forming a true belief. Can one go wrong in this domain? Certainly. An agent learning a new language will struggle to form true beliefs reliably in that language– something very familiar to those who have attempted to learn a new language. But with enough training, one can become sufficiently competent and use their comprehension abilities to form true beliefs reliably, upon receipt of veridical testimony in that language. And so, if we see comprehension as the epistemically significant competence through which we "get things right", then the anti-monitoring theorist can once more label monitoring as redundant, since we can make sense of reliable testimonial belief formation without appealing to it. And this puts us back where we were. Virtue-reliabilism does not rescue reliabilist MVs.

[5] Burge adds that a capacity to understand not only content, but the *force* of content is necessary for our entitlement to accept what we are told (ibid.). That is, we must have the capacity to understand that an utterance is an assertion and not, say, a joke. Burge defends this position further in his 1999 essay "Comprehension and Interpretation".

Let us now reconsider the case of the strikingly gullible Polyanna, as discussed in § 2. One might forward that my arguments permit a rather counterintuitive conclusion: Polyanna *can* acquire testimonial knowledge. And insofar as my arguments allow this, I must be incorrect. But there are a few things to note here. Firstly, if one supports reliabilism, then ascribing knowledge to Polyanna need not be obviously problematic, depending on the environment she inhabits. If she inhabits an environment with a very high base rate of speaker-honesty, then the process responsible for outputting her testimonial beliefs is a reliable one *in that environment*. As such, it is not clear why one should reject, *on purely reliabilist grounds*, that Polyanna is a knower notwithstanding its perceived counterintuitiveness.

However, if one is convinced that Polyanna cannot gain testimonial knowledge, then my arguments force one to look beyond issues relating to reliability. The intuition that Polyanna fails to acquire testimonial knowledge cannot depend on viewing her as forming testimonial beliefs unreliably, since we must concede that Polyanna forms testimonial beliefs reliably if she inhabits an environment with a very high base rate of speaker-honesty. Perhaps the issue is instead that Polyanna fails some kind of *rationality* condition, provided that this is understood as something other than a mere tendency towards true belief. If, as we often see, rationality is understood as an internalist constraint on justified belief, then this would seem to work.[6] But if Polyanna's epistemic fault is that she forms beliefs irrationally, she fails not only to attain testimonial knowledge but justified testimonial belief. As such, defending the claim that Polyanna cannot acquire testimonial knowledge due to her failing a rationality condition implies implicit support of an internalist stance with respect to epistemic justification.

I wish to note that it is entirely compatible with the arguments provided in this paper to maintain both that reliability is epistemically significant *and* that monitoring is necessary for testimonial knowledge. As I have shown, however, one's defence of that monitoring condition cannot rest solely on reliabilist grounds. Indeed, I think that the door is left open for WM on the grounds that it is required for agents to meet a rationality condition on knowledge, where rationality consists in, say, the coherence of our doxastic states (Sosa, 1985). But the claim that monitoring is needed because it establishes reliability is precisely what I reject. As I see it, if you prize reliabilism, you should look to AMVs, and if you prize MVs, you should look beyond reliabilism.

If one wishes to defend MVs, one will have to appeal to something *beyond* reliability. One might argue, for example, that, reliability aside, acceptance-without-monitoring constitutes an *epistemic vice* insofar as it is associated with a defective inquiry (Hookway, 2003: 118), (Baehr, 2011: 18), or that it indicates an inadequate motivation to reach the truth (Montmarquet, 1992: 336). Say one opts for this route. In doing so, one is committed to saying that the justificatory status of testimonial beliefs is *not* determined solely by the reliability of the processes responsible for outputting it. Indeed, one could opt for this route while maintaining that acceptance-with-monitoring and acceptance-without-monitoring are as reliable as each other. Advocates of this route will instead hold that, for one's belief that p to be a candidate for knowledge, one must reach the belief that p in an "epistemically virtuous man-

[6] See Feldman and Conee (1985), Sosa (1985), Foley (1987).

ner" (Zagzebski, 1996). Of course, the monitoring theorist who defends this virtue-theoretic position owes us an explanation of what that virtuous manner is and why this excludes acceptance-without-monitoring. But I do not wish to suggest that one cannot make such an argument, nor that MVs are doomed to failure. Instead, I have argued that MVs cannot be defended *on reliabilist grounds*. For the reliabilist monitoring theorist, this presents a dichotomy: they can prioritise MVs over reliabilism, looking beyond reliability to defend monitoring as necessary for testimonial knowledge; or they can prioritise reliabilism over MVs, arguing that acceptance-without-monitoring constitutes a reliable process. Either way requires a major revision to one's epistemology of testimony.

Acknowledgements I am indebted to Paul Faulkner, Jack Lyons, Viktoriia Kononova and an anonymous reviewer for their extensive comments.

Funding The author received no funding to assist with the preparation of this manuscript.

Data availability Not applicable.

Declarations

Conflict of interests The author has no relevant financial or non-financial interests to disclose.

References

Baehr, J. (2011). *The inquiring mind: On intellectual virtues & virtue epistemology*. Oxford University Press.

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

Burge, T. (1993). Content preservation. *Philosophical Issues*, *6*, 271–300.

Burge, T. (1997). Interlocution, Perception, and Memory, in: T. Burge (Ed.), *Cognition Through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection: Philosophical Essays, Volume 3* (2013), Oxford: Oxford University Press, 285–305.

Burge, T. (1999). Comprehension and Interpretation, in: T. Burge (Ed.), *Cognition Through Understanding: Self-Knowledge, Interlocution, Reasoning, Reflection: Philosophical Essays, Volume 3* (2013), Oxford University Press, 342–361.

Dunn, J., & Ahlstrom-Vij, K. (2017). Is reliabilism a form of consequentialism? *American Philosophical Quarterly*, *54*(2), 183–194.

Elster, J. (1989). Social norms and economic theory. *The Journal of Economic Perspectives*, *3*(4), 99–117.

Faulkner, P. (2010). Norms of Trust, in: A. Haddock, A. Millar, and D. Pritchard (Eds.), *Social Epistemology* (2010), Oxford University Press, 129–147.

Feldman, R., & Conee, E. (1985). Evidentialism. *Philosophical Studies*, *48*(1), 15–34.
Foley, R. (1987). *The theory of epistemic rationality*. Harvard University Press.
Fricker, E. (1994). Against gullibility. In B. K. Matilal, & A. Chakrabarti (Eds.), *Knowing from words*. Kluwer Academic, 125–61.
Fricker, E. (2017). Inference to the best explanation and the receipt of testimony. In K. McCain, & T. Boston (Eds.), *Best explanations: New essays on inference to the best explanation* (pp. 262–294). Oxford University Press.
Goldberg, S. (2007). *Anti-Individualism: Mind and language, knowledge and justification*. Cambridge University Press.
Goldberg, S. (2010). *Relying on others: An essay in epistemology*. Oxford University Press.
Goldman, A. (1979). What Is Justified Belief? in A. Goldman (Ed.) (2012), *Reliabilism and Contemporary Epistemology: Essays*, Oxford: Oxford University Press, 29–48.
Graham, P. J. (2010). Testimonial entitlement and the function of comprehension. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Social epistemology* (pp. 148–174). Oxford University Press.
Graham, P. J. (2015). Epistemic normativity and social norms. In D. K. Henderson, & J. Greco (Eds.), *Epistemic evaluation: Purposeful epistemology* (pp. 246–273). Oxford University Press.
Greco, J. (1993). Virtues and vices of virtue epistemology'. *Canadian Journal of Philosophy*, *23*(3), 413–432.
Greco, J. (2003). Knowledge as credit for true belief. In M. DePaul, & L. Zagzebski (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology* (pp. 111–134). Oxford University Press.
Hookway, C. (2003). How to be a virtue epistemologist. In M. DePaul, & L. Zagzebski (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology* (pp. 183–202). Oxford University Press.
Levine, T., Park, H., & McCornack, S. (1999). Accuracy in detecting truths and lies: Documenting the veracity effect. *Communication Monographs*, *66*, 125–144.
Lyons, J. (2013). Should reliabilists be worried about demon worlds? *Philosophy and Phenomenological Research*, *86*(1), 1–40.
Michaelian, K. (2010). In defence of gullibility: The epistemology of testimony and the psychology of deception detection. *Synthese*, *176*, 399–427.
Montmarquet, J. (1992). Epistemic virtue and doxastic responsibility. *American Philosophical Quarterly*, *29*(4), 331–341.
Plantinga, A. (1986). Epistemic justification. *Noûs*, *20*(1), 3–18.
Plantinga, A. (1988). Positive epistemic status and proper function. *Philosophical Perspectives*, *2*, 1–50.
Pritchard, D. (2010). Anti-Luck virtue epistemology. In D. Pritchard, A. Millar, & A. Haddock (Eds.), *The nature and value of knowledge: Three investigations* (pp. 48–65). Oxford University Press.
Simion, M. (2021). Testimonial contractarianism: A knowledge-first social epistemology. *Noûs*, *55*(4), 891–916.
Sosa, E. (1985). The coherence of virtue and the virtue of coherence: Justification in epistemology. *Synthese*, *64*(1), 3–28.
Sosa, E. (1991). *Knowledge in perspective: Selected essays in epistemology*. Cambridge University Press.
Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge*. Oxford University Press.
Sosa, E. (2015). *Judgment and agency*. Oxford University Press.
Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393.
Vrij, A. (2008). *Detecting Lies and deceit: Pitfalls and opportunities* (2nd ed.). Wiley.
Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.