
Book Section:
Bennett, C. (2021) Eyewitness testimony, the misinformation effect and reasonable doubt. In: Hoskins, Z. and Robson, J., (eds.) The Social Epistemology of Legal Trials. Routledge (Taylor & Francis), pp. 30-45. ISBN: 9780367245535.

https://doi.org/10.4324/9780429283123-2

Eyewitness Testimony, the Misinformation Effect, and Reasonable Doubt

Christopher Bennett

c.bennett@sheffield.ac.uk

ABSTRACT

In a recent paper, Katherine Puddifoot has argued that jurors should be given information about the misinformation effect in order to preserve the useful role that eyewitness testimony can sometimes play in criminal trials, while mitigating the distortions to which the misinformation effect might give rise, while. I will argue that her strategy, while promising, does not go far enough. Because it involves agreeing that the misinformation effect will foreseeably distort some eyewitness testimony, Puddifoot's strategy cannot answer the charge that, given the high standard of proof in criminal trials (beyond reasonable doubt), it would be better to disallow convictions in all cases in which eyewitness testimony is central or decisive. To consider whether such disallowing would be an appropriate, proportionate course of action, I will claim, we need to get clearer about issues concerning the appropriate standard of proof in criminal trials, and about the values in play that should help us interpret that standard.

1. Introduction

Sometimes we can have memories suggested to us about events that did not obtain. For instance, Elizabeth Loftus and colleagues presented experimental subjects with images of a car crash, and found that subjects who were asked how fast the cars were travelling when they 'smashed' into each other were more likely falsely to report that there was broken glass in the images than those who were asked how fast the cars were travelling when they 'hit' each other (Loftus 2005; Zaragoza et al 2007). Following the usage of Loftus and colleagues, I will talk about these cases of suggestion as involving 'the misinformation effect.' In this paper I am particularly concerned with whether the existence of the misinformation effect reduces or removes the justifiability of the

widespread practice of relying on eyewitness testimony in criminal trials. I will distinguish two possible lines of response. The first is a mitigating response. It has been argued, for instance, that the way to address the issue is that to ensure that jurors are presented with evidence about existence of the misinformation effect, and about the possibility that some witness testimony might be contaminated by it. However, I will contrast this mitigating response with a second line of response that would be more drastic. To proponents of this second line of response, mere mitigation fails to remove the real source of the problem, which is that it would be incompatible with the criminal law's standard of proof of beyond reasonable doubt (BARD) to leave the matter to juries. Given a commitment to BARD, this more drastic response argues that we should go further, disallowing convictions in all cases in which eyewitness testimony is central or decisive. For instance, if a particular witness identification of a suspect at a crime scene is a key piece of evidence in the prosecution case in a given criminal trial, and where that evidence is not subject to corroboration by source of evidence, this more drastic response argues that the case should not go forward. Those who are concerned to say how the criminal justice system should deal with evidence of the misinformation effect have to decide which of these lines of response – merely mitigating or more drastic – to take.

In a recent paper, Katherine Puddifoot has argued that jurors should be given information about the misinformation effect in order to mitigate the distortions to which it might give rise, while preserving the useful role that eyewitness testimony can sometimes play in criminal trials. Indeed, her claim is that, if jurors are to receive advice about the existence of the misinformation effect, this advice should be altered in order to prevent jurors from overcompensating and ruling out eyewitness evidence that may

be perfectly reliable and accurate. Puddifoot's strategy can be seen to be a variant of the mitigating line of response (Puddifoot 2020). However, Puddifoot does not herself raise the question of whether it would be better, given BARD, to rule out cases in which eyewitness evidence is decisive and uncorroborated. Given this, however, I will argue that her strategy must be incomplete since it assumes the validity of the mitigating line of response and does not explain why we it would not be more in keeping with the strictures of BARD to take the more drastic line of response in all cases in which eyewitness testimony is central or decisive. To consider whether allowing or disallowing eyewitness testimony in cases in which it is decisive would be the appropriate, proportionate course of action, I will claim, we need to get clearer about issues that Puddifoot does not broach, and which concern the appropriate standard of proof in criminal trials, the proper interpretation of BARD, and the values that should help us interpret that standard. I argue that once we look at these more foundational issues we will see that the drastic response would in fact be out of proportion and that the mitigating response defended by Puddifoot is the correct one.

2. Puddifoot on the 'overcritical juror'

Puddifoot draws on an impressive array of evidence in social and cognitive psychology to argue that a widely recognised source of memory distortion may in fact lead us to be disproportionately sceptical of eyewitness testimony. Puddifoot argues that there is good reason to think that human beings – eyewitnesses to crime included – can be subject to the *misinformation effect*. The misinformation effect, as Puddifoot explains it, occurs as a result of misinformation provided to a person after an event, and has the effect of introducing non-existent features of that event that the person incorporates into their memories and then reports as being genuine features of the event. Puddifoot

describes two ways in which the misinformation can come about (Puddifoot 2020, p. 3): first of all, through suggestive lines of questioning about the event (for instance, 'how fast was the white sports car going when it passed the barn while travelling along the country road?' – when in fact there was no barn); and secondly, through a wider social contamination effect that can occur when we are given misinformation by 'unfamiliar social peers.' To transpose this to the case of eyewitnesses, then, we could imagine that they could be subject to the misinformation effect as a result of suggestive questioning by police officers, or as a result of discussing the events with other eyewitnesses. However, this can then lead eyewitnesses to give testimony that, at least in details, can be demonstrated to be false at cross-examination. For instance:

> 'An eyewitness describes a car travelling at a specific speed but CCTV evidence shows that the car was travelling more quickly or slowly. An eyewitness claims that a car involved in an accident passed a yield sign when there was actually only a stop sign at the scene. An eyewitness claims that the person who committed a crime was wearing glasses but photographic evidence suggests that they were not.' (Puddifoot 2020, p. 6)

In this case it can appear rational for the jury to conclude that the eyewitness is untrustworthy. The witnesses can appear either to be deceptive, or lacking in competence, but either way the jury's trust in what they have to say is likely to be undermined. However, Puddifoot argues that this conclusion would be too hasty. For a start, if memory distortion is brought about by the misinformation effect, it is clearly not the result of deliberate deception. Furthermore, Puddifoot argues that being subject to the misinformation effect may not be a sign of being an unreliable witness at all. She

claims that the psychological evidence reveals the misinformation effect to be the result of the normal workings of normally reliable human cognitive mechanisms, and in particular our tendency to incorporate information acquired after an event into our memories of that event. This tendency towards incorporation is what allows the misinformation effect to come about, since it seems that it is this tendency that primes us to accept suggestion. However, Puddifoot argues that this tendency to incorporation is part and parcel of our ability *accurately* to update memories in the light of new information, an ability that we need to have if our memories are, like other beliefs, to keep track of our evolving epistemic situation. And if being subject to the misinformation effect is a side-effect of epistemic virtue – that is, of the normal workings of an epistemic mechanism that is normally beneficial and reliable – then it should not discredit the witness:

> 'if leading theories in cognitive science and contemporary philosophy of memory are correct, the errors that occur due to the misinformation effect are the result of a feature of human cognitive systems which can bring substantial epistemic benefits. Moreover, the epistemic benefits gained through this feature or these features of human cognition increase the chance of any person being a good eyewitness. This means that errors in testimony can be a sign of the ordinary operation of the cognitive mechanisms that make human beings able to be good eyewitnesses.' (Puddifoot 2020, p. 11)

On the constructivist picture of memory with which Puddifoot is working, memory stores traces of information about an event (possibly from various sources), which are then combined when 'recall' is demanded (Puddifoot 2020, p. 8, p. 12). This is the

normal – and normally effective – working of the mechanism of memory; but it means that it is a structural risk of the mechanism (compatible with it being epistemically virtuous) that some of the information that may have come from testimony received after the event is itself unreliable. Once that testimony is accepted by the subject, and the mechanism then plays its normal role in updating the relevant memories, those memories will be 'suggested' rather than accurate. But the key point is that this is quite compatible with the mechanism of memory itself working perfectly reliably. Therefore Puddifoot's point is that the witness should not too quickly be discredited even if a particular piece of their evidence does turn out to have been subject to the misinformation effect.

A further complication that Puddifoot draws from recent work about memory, and which can add to our awareness of our susceptibility to the misinformation effect, is that we have gist memories and verbatim memories (Puddifoot 2020, p. 9). While the latter store precise details of events, the former store broad features of events where the details have to be reconstructed at the time of recall. One of the ways in which we are susceptible to memory distortion through the misinformation effect is that our verbatim memories are more likely to fade across time, and gist memories remain. This is why we can be vulnerable to the phenomenon of 'remembering' false details that have been suggested to us – and Puddifoot supplies a useful list of further factors that might increase the likelihood of such suggestion, such as the amount of time that has passed, or whether the accent of the person giving the false information conveys that they are powerful and socially attractive (Puddifoot 2020, p. 10). Nevertheless, Puddifoot points out, this again shows that being subject to the misinformation effect is quite compatible with epistemic or intellectual virtue – that is, being an acute, attentive, critical and

unprejudiced observer, and hence everything that one might want in an eyewitness – and that the misinformation effect is quite compatible with our having a good supply of 'gist' memories of the event (Puddifoot 2020, p. 10).

As a result, Puddifoot claims, the guidance given to jurors in cases about the ways in which eyewitnesses can be wrong needs to be altered so as not to neglect the important ways in which those eyewitnesses can be right even when subject to the misinformation effect. Simply introducing warnings about the misinformation effect, without the nuanced analysis of its significance that Puddifoot provides, would lead to jurors being *overcritical* of eyewitness testimony, with the result that an important source of evidence for the prosecution might be lost.

> 'Numerous authors have argued that jurors should be made aware of these findings so that they can lower the credence that they give to eyewitness testimony to reflect its susceptibility to error due to the misinformation effect. What the current discussion shows is that the psychological findings can be used to achieve the opposite goal of ensuring that the credence given to individual pieces of eyewitness testimony is not lowered inappropriately.' (Puddifoot 2020, p. 20)

To sum up the thrust of Puddifoot's argument as I have summarised it here, her strategy is not to deny that the misinformation effect exists, but rather to point out that susceptibility to suggested memories may be a side-effect of the normal, and normally accurate and beneficial, workings of the mechanisms involved in our ability to store and recall memories. Her conclusion is therefore that we should not discredit the witness as

a reliable source of information in virtue of the fact that they are found to have been subject to the misinformation effect; and that advice given to jurors should be revised in order to reflect an eyewitness's likely reliability.

However, if we now ask how Puddifoot's conclusion should inform our thinking about whether the mitigating and drastic lines is most appropriate given the standard of proof in criminal trials, we might worry that she has not given us enough to rule out the drastic response. Her line of argument takes it for granted that some version of the mitigating response is viable. However, because she allows that the misinformation effect does take place, it is an implication of her argument that any system that does not disallow eyewitness evidence in trials in which it is decisive and uncorroborated would, given the impossibility of insulating witnesses from suggestion, foreseeably give rise to cases of mistaken identity that result in wrongful conviction. Given that we have grounds to register this implication in advance of knowing the details of any particular trials, the proponent of the drastic response can argue that, if we are genuinely committed to allowing convictions only when there is no reasonable doubt, we should act in advance to rule out the possibility of such wrongful conviction by disallowing any such trial. This is the argument that I will explore and respond to in the rest of this chapter.

Before I get on to that, however, I should note that in one place Puddifoot does suggest a stronger line of response. While the misinformation effect seems puzzling and devastating because it involves mistaking memories of testimony for memories of direct experience, Puddifoot claims that the effect may not be as serious as some of the initial research suggested. To defend this claim she points to evidence regarding our source-

monitoring capacities: that is, the capacity we have of identifying the source from which we derived the memory in question. She claims that the evidence shows that we are generally reliable in identifying the type of source from which we have derived the memory (Puddifoot 2020, pp. 10-11). Source-monitoring 'involves comparing the experiential characteristics associated with the representations to expectations of what the characteristics would be like if they were produced by, for example, experience rather than dreaming' (Puddifoot 2020, p. 11). So the idea is that experiencing something as veridical is different from experiencing it as a confection. And presumably the point is that, because source-monitoring comes about through the representations bearing something akin to a stamp of origin, these experiential differences reliably align with the distinction between veridical memories – those caused by actual events – and false or confected memories. Thus these experiential differences between the representations 'ensure that people are often good at identifying their source through this comparison' (Puddifoot 2020, p. 11).

We might draw two conclusions from this point about our adeptness at 'source-monitoring.' One is that this is further evidence that, insofar as a person is subject to the misinformation effect this is likely to be an isolated error and 'does not mean that she is likely to make numerous other errors' (Puddifoot 2020, p. 11). This is again to suggest that susceptibility to the misinformation effect is not incompatible with overall good epistemic functioning and reliability. However, if this was the only implication of the point about source-monitoring then it would not challenge our concern that there is still reasonable doubt about whether the misinformation effect has taken place in any given case. Another implication that Puddifoot might draw, though, is that, because the research suggests that we are in fact often good at identifying the source of our

memories, the misinformation effect is likely to be much less prevalent than some have worried it was. If the evidence is really that our actual susceptibility to the misinformation effect is in fact vanishingly rare then we might think that the drastic response of ruling out eyewitness testimony is not a plausible one to begin with, and hence not really a serious challenge to any variant of the mitigating response. In which case the wider investigation of the proper interpretation of BARD that I have proposed would be unnecessary.

However, I do not think that we can conclude that the misinformation effect *is* vanishingly rare. Now obviously, it is a complex empirical matter how prevalent the effect is, and one that merits further investigation. However, one of the reasons that Loftus's research has caught the imagination, not just of scholars, but of the wider public is that the research chimes with a phenomenon to which we intuitively recognise ourselves to be at least sometimes subject. Indeed, intuitive evidence for the unreliability of memory, and the possibility of motivated or suggested memories is a common feature of everyday discourse ('her memory is playing tricks on her again …'). Furthermore, Puddifoot herself admits that our source-monitoring is not infallible. She agrees that the misinformation effect takes place and that 'the fallibility of the process [of source-monitoring] explains how the misinformation effect occurs': hence the reality of the misinformation effect shows that our source-monitoring can go wrong. It seems to be agreed by Puddifoot, in other words, that Loftus and colleagues have shown, in an experimental setting, that in at least some apparently revealing cases we are unable to discriminate between memories that are veridical and memories that are suggested. This is enough, I submit, to raise a case that has to be answered about whether there is inevitably reasonable and hence problematic doubt in any trial in which eyewitness

evidence is both decisive for the prosecution case and uncorroborated by an independent source.

3. Do Puddifoot's claims remove grounds for reasonable doubt?

While granting the point that our source-monitoring is generally reliable, there seems to be little denying, in Puddifoot's account, that individual pieces of remembered information are susceptible to corruption through suggestion. Given that our source-monitoring is fallible, and that risk factors such as being subject to potentially suggestive questioning, and the passing of long periods of time between the event and its being recalled, are often a characteristic feature of criminal trials, it seems that jurors still have a job to do if they want to make a responsible assessment of the significance of the eyewitness's testimony. Puddifoot acknowledges as much when she explains how jurors should reflect on the modified guidance about the misinformation effect that she recommends.

> 'Once these details were available to jurors, they could apply their new knowledge about the misinformation effect to evaluate whether an individual is likely to have undergone the effect, and whether any individual piece of evidence is likely to have been distorted as a result. For example, a juror informed by the psychological findings might consider whether an eyewitness could have been repeatedly questioned about a particular detail of an event and consequently misremembered the detail due to the misinformation effect. Then the juror can consider whether the eyewitness is likely to have been subjected to similar questioning about other details about which she might testify.' (Puddifoot 2020, p. 20)

However, it seems that the proponent of the more drastic line of response can now argue that, for all Puddifoot's efforts to rehabilitate eyewitness testimony, she still leaves room for reasonable doubt about particular pieces of eyewitness testimony. This seems particularly undeniable in any case in which a particular piece of evidence from a particular eyewitness would be central or decisive. Furthermore, in order to show that conviction would not be beyond reasonable doubt, it would not be necessary to show that the witness had *in fact* been subject to the misinformation effect. Of course, one strategy for discrediting an eyewitness might be to attempt to show that this particular person in fact has been subject to the misinformation effect, by showing that they were in fact subject to suggestive questioning, or showing that they had in fact participated in suggestive discussions with other eyewitnesses. However, for this to be demonstrated would be very costly: records of police questioning or of the conversations actually held between eyewitnesses would need to be consulted, and then it would have to be shown that something that might be misinformation was actually introduced at these points. Another, more likely way in which the discrediting strategy might be carried out would involve, not so much showing that suggestion had in fact taken place, but that it is *sufficiently probable* that it had taken place to give rise to reasonable doubt. For instance, counsel might be able to show that an eyewitness had had (or that it can't be ruled out that they had) conversations with other eyewitnesses (perhaps without being able to specify their content), and would then be able to argue that, on the basis that the eyewitness might now be affected by the misinformation effect, the testimony was now not to be relied on. Given that this strategy would be available, and apparently justifiable, in almost any criminal trial in which evidence from a single eyewitness is decisive, this seems to give us a strong argument in favour of distrusting – and hence

pre-emptively disallowing – convictions in such cases as incompatible with excluding reasonable doubt. For instance, the authors of the 1976 Devlin report on evidence of identification in criminal cases concluded:

> 'We do however wish to ensure that in ordinary cases prosecutions are not brought on eye-witness evidence only, and if brought, they will fail. We think that they ought to fail, since in our opinion it is only in exceptional cases that identification evidence is sufficiently reliable to exclude a reasonable doubt about guilt.' (Devlin Report 1976: pp. 149-150)

Thus if this more drastic strategy is the one that appears to be called for by the principles underlying BARD, it becomes a bit less clear how Puddifoot's arguments are meant to help to restore credibility to the eyewitness. Puddifoot agrees that the eyewitness may be unreliable with respect to some pieces of testimony, even though they may be highly reliable in regard to others. However, unless we are able to pinpoint exactly which testimony is reliable and which is not (as we would if the counsel were following the first discrediting strategy, the one that seems less likely), it seems as though, for all Puddifoot's painstaking review of the relevant research, we are still left with the conclusion that we do not know which pieces of the eyewitness's testimony we can rely on and which we cannot. And if we know that there is some reasonable chance that suggestion might have taken place, there is room for reasonable doubt, and that is all that is necessary. The mitigating line of response is incompatible with BARD.

4. The epistemic risk response

I will now offer an argument defending the mitigating strategy against the drastic strategy. I will claim that it would be a disproportionate response to the misinformation effect to rule out conviction in any case in which uncorroborated eyewitness evidence plays a decisive role. I will argue that the ends of justice can sometimes be served by conviction in a case in which the conclusion that the defendant is guilty is fallible. Thus I will point to a way of defending Puddifoot's conclusion that we should allow juries to weigh the risks of error, rather than disallowing such cases in advance. I suggest that we start with what I will call the *epistemic risk response*.[1]

To see how the epistemic risk response might go, we will look at a parallel argument in epistemology. We are interested in how to interpret and understand the implications of BARD; and as we have interpreted it thus far, BARD seems to undermine any eyewitness evidence about which there is foreseeably the chance of error. If the mitigating strategy has any chance of success then that interpretation has to be shown to be wrong. The parallel argument that we will look at from epistemology is an argument that seeks to rescue ordinary knowledge-claims from the corrosive effects of scepticism. The sceptic begins by pointing out the fact that we know that we are not epistemically infallible, and hence that we foreseeably have false beliefs; then claims that, in any given situation, we cannot tell whether a belief that we rely on in an inference or as the basis of action is one of those false beliefs or whether it is true; and then draws the conclusion that, since we cannot tell which of our beliefs are true and

[1] I have drawn the epistemic risk response from two broadly pragmatist sources. Firstly, the 'relevant alternatives' account of knowledge as canvassed in Goldman (1976). And secondly, the view that the importance of 'what is at stake' in a situation properly influences knowledge attributions: for some discussion, see DeRose (1992), Fantl and McGrath (2002) and Stanley (2007). The discussion of these views in epistemology focuses on when we can say that a person 'knows;' although in the text I also talk about 'knowledge,' something more pragmatic is enough for our purposes, such as when we are sufficiently sure of a proposition to be justified in taking it as the basis for action.

which are false, we cannot be in a position to have knowledge about the content of any of our beliefs (or rely on them in inferences or as the basis for action). This sceptical argument is a close parallel to the drastic argument from BARD. And just as many think that scepticism is untenable and that its assumptions must be in some way problematic, so proponents of the mitigating strategy might argue the same about the drastic strategy. My proposal is that we look at anti-sceptical strategies in epistemology in order to see what alternatives there might be to the drastic strategy.

To start with, consider that the considerations about the fallibility of memory that we have been looking at here might be used to ground a general scepticism. The argument against Puddifoot that we have canvassed need not deny that many memories are veridical. But it does point out that we are not in a position, in the case of any memory that may potentially have been subject to suggestion, to say which are veridical and which are suggested. The only distinguishing marks of veridicality, from the point of view of the person remembering, are phenomenal, and, despite our source-monitoring abilities the research on the misinformation effect seems to show that suggested memories can be phenomenally indistinguishable from veridical ones. And if that is true then it looks as though it will *always* be an open question, with regard to *any* memory, whether it was suggested. And this in turn might seem to lead us to the devastating sceptical conclusion that we can *never* really know whether any memories that might have been subject to suggestion are trustworthy. Thus we cannot really be said to know whether the facts that our memories appear to present to us as having obtained really did obtain.

Furthermore, this has general implications for the question of what we can learn from others. As is widely remarked, one of our major sources of knowledge is what others tell us. Many theorists think that testimony allows, in many cases, the *transfer of knowledge*. That is, testimony operates to transfer epistemic warrant in such a way that the person getting told can have as much claim to know what they have been told as does the teller. However, now consider that much of that testimony concerns events in the past. Such testimony is therefore based on the memories that the testifier has of those past events. If we have now been forced to conclude that the testifier has no knowledge of those events in any case in which they have not been able to rule out the influence of suggestion then it seems as though the recipient of their testimony also cannot be said to know.

However, according to what I will call the *epistemic risk response*, this general scepticism about the epistemology of memory and testimony is unwarranted. The epistemic risk response attacks the idea that we need certainty, or infallibility, in order to have knowledge, or to be able to rely on beliefs in inferences and as a basis for action. The epistemic risk response has two parts: first of all, it claims that in order to have knowledge one would need to be able to show, not that all logically possible alternatives do not obtain, but only that *relevant* alternatives do not obtain; and secondly, it claims that the way to show which alternatives are relevant is by appeal to a notion of risk. Let us describe the epistemic risk response in more detail.

In order to have infallibility we would need to rule out all logically incompatible alternatives before we can be said to have trustworthy belief, or indeed knowledge. If we want to assert that certain facts XYZ obtain; and if it is the case that, had some

alternative scenario ABC obtained, XYZ could not have obtained; then it looks as though we need to be in a position to show that ABC did not obtain before we are able to claim that XYZ did. Furthermore, if we allow that some alternative scenario *logically might* have obtained then it seems as though we cannot be entirely certain that it did not. We can see that this idea figured in the above sceptical argument in the shape of the claim that, unless we can rule out the possibility that a given memory was suggested then we don't know that that memory is veridical, and hence we don't know that the events that the memory represents as having taken place really did take place. However, this premise is rejected by the epistemic risk response in any form that would generate a general scepticism. The epistemic risk response agrees that, if we are after certainty then we do indeed need to know that no logically possible incompatible alternative obtains, but it claims that it can be sufficient for knowledge – given the practical role that knowledge-claims play as a basis for inference and action – that we have something less than absolute certainty. We need to rule out alternatives, but rather than having to rule out absolutely everything that may have been the case, we can introduce some pragmatism into our search for trustworthy belief and knowledge. What we need to be able to rule out in order to ground a claim of knowledge are, not all logically possible alternatives, but rather all *relevant* alternatives.

This is an important claim in the argument against scepticism since it allows us to say that e.g. wild speculation about being a brain in a vat is not a relevant alternative, and that we do not need to prove that this scenario does not obtain before we are able to say e.g. that I can see a hand before me. We clearly do not tend to think that we need to rule out brain-in-the-vat alternatives before we can rely on the existence of our hands (and of the external world more generally), even though a) there seems no way of showing

conclusively that the brain-in-the-vat scenario does not obtain, and b) there might be situations such as having a hallucination that would make us reasonably doubt whether our hands before us really are as we think we see them.

Nevertheless, even if this strategy is promising, the notion of 'relevance' clearly needs to be pinned down a bit if this argument is to do any real work. In particular, if we are to block scepticism about memory, we would need to know why the possibility of the misinformation effect is not a relevant possibility in this sense. According to the epistemic risk response, what makes an alternative scenario relevant is that it represents a genuine risk to our being able to rely on a given belief in making inferences and as the basis of action. Epistemic risk, like other types of risk, comprises two factors, probability and stakes. According to the epistemic risk response, then, knowledge is not an all-or-nothing matter of certainty, but rather comes in degrees of credence. Assigning credence is governed by norms, and the degree of credence (or assent) that we are warranted in giving to a proposition is dictated by these two factors: first of all, the probability that we are warranted, given the evidence available to us, in assigning to its being true; and secondly, what is at stake in getting it right, that is, how significant the consequences would be of getting it wrong. Thus the importance of ruling out alternative incompatible explanatory scenarios itself comes in degrees: a failure to rule out some alternative explanatory scenario weakens our warranted credence in proportion to the probability that that incompatible scenario obtains, and the more it would matter if I got it wrong. In other words, we are often right to entertain doubts about whether we really know something, and to attempt to silence those doubts by gaining more evidence, but those doubts should be reasonable, and reasonableness,

according to the epistemic risk strategy should be thought of in terms of probability and stakes.

Applying the epistemic risk response to the misinformation effect, we can now say the following. While it may be true that the bare possibility that suggestion has happened should make us judge that a given memory is not *entirely* trustworthy, and hence that we cannot be *entirely* sure that the events the memory depicts as having obtained really did obtain, this kind of absolute certainty need not be required for warranted belief or even knowledge. What we really need to attend to is the probability that some alternative explanatory scenario obtained – the probability, that is, that there was suggestion – what our interests are in being able to make the judgements in question, and why it would matter if we got it wrong. This of course means that we need to make some estimate of how likely it is, in any particular case, that a witness's testimony has been subject to the misinformation effect. But we also need to think, once we have some notion of the probability, why we want to be able to make judgements of culpability in the first place, and how much it matters whether we get it wrong, and why. When we are thinking about stakes in the criminal trial then we are presumably thinking that we have an interest in being able to make judgements of culpability because of the reasons we have for trying and punishing those suspected or convicted of criminality (be it deterrent or retributive reasons, etc.), and at least two major costs of getting it wrong: the cost of failing to do justice by letting a guilty person go unconvicted; and the cost of sanctioning, and censuring, an undeserving person. It is hard to see how we can make an informed judgement about the level of credence that is sufficient for judgments of culpability without broaching, not just questions of probability, but also questions of what is at stake in getting it right.

5. A proportionate response to epistemic risk?

Putting this otherwise, we need to look at the importance of the goals that are furthered or realised by deserved conviction, how probable it is that false negatives and false positives will occur, and how bad that would be (and for whom). The answers to these questions might be very different across different criminal justice systems, even if those systems share a broad conception of their general justifying aim. If the consequences of conviction can be analysed into a censuring component, a sanctioning component, and a further consequences component, what is at stake will surely depend on such things as: how stigmatising the censure is and how easily a person convicted of crime can achieve redemption and regain respect in their own eyes and the eyes of the community; what kinds of punishments those convicted are subject to, how intrinsically unpleasant they are, and how long they last for; how much damage conviction does to the wider life prospects of an individual, and of those of their dependents and communities; how effective punishment is in contributing instrumentally to goals such as reducing crime; how important are any non-instrumental goals of punishment, such as vindicating victims, dissociating the political community from wrongdoing, taking wrongdoing seriously, etc.. There might also be a general question of whether the criminal justice system exacerbates or ameliorates social injustice more broadly. If, as W. H. Simon suggests, the rational approach for criminal defence lawyers, given the inhumaneness, disproportionate severity and injustice of the current system, is to seek to get their clients off at all costs, no matter their substantive guilt or innocence, then we should favour a very high bar for conviction and should require something approaching certainty before consigning any person to such a fate (Simon 1993: pp. 1722-3).

Perhaps, though, things are, while seriously problematic, not quite so bad that we should always favour that strategy that promises to minimise the number of convictions. Perhaps also we could employ a degree of idealisation, and ask about the adequate way to interpret BARD in a system in which censure is not too stigmatising, where real possibilities for social reintegration exist and are widely taken up, and where prisons are humane and imprisonment used sparingly. In such a situation, where there is either clear evidence that punishment reduces crime, or the non-instrumental goals of punishment can be argued to be of sufficient importance, would it be disproportionate to rule out eyewitness evidence altogether in cases in which it is decisive and uncorroborated?

The crucial question is whether there are categories of crime that are likely to be made unprosecutable by virtue of such a restriction, and whether a proportionate approach to fairness to the defendant nevertheless requires us to accept that restriction. For instance, many cases of sexual assault take place in circumstances in which the attacker has purposefully isolated the victim, or has attacked them when they are isolated. If we accepted the drastic approach, we would have to accept that prosecutions for such cases would be ruled out or very unlikely to succeed. To my mind that would be uncomfortably close to allowing perpetrators of sexual attacks impunity. While we should recognise that moral constraints may regrettably but categorically limit our ability to prevent or prosecute crimes that we would ideally wish to, we should also bear in mind the pattern of costs and benefits that stems from our adoption of a set of constraints as authoritative, in particular where the costs fall disproportionately on those who are already vulnerable. Several decades of feminism do not yet appear to have done much to dislodge the perception among many men that sexual access is their

default prerogative. This perception seems to be one of the most widespread, most stubborn, and most corrosive impediments to guaranteeing our fellow citizens a decent life free from fear and living together as equals. It is for this reason that Claudia Card has called rape a 'terrorist institution' (Card 1991). If we recognise the importance of taking action against sexual attacks through the criminal justice system then I think we cannot accept the drastic solution if it implies that there would only rarely be successful prosecutions for sexual attacks.

Given these considerations about the need to keep open the possibility of prosecution for sexual attacks and other predatory behaviour, I submit that the epistemic risk response is a promising way to defend the mitigating strategy. In determining what is at stake in getting it right, we have to look at the fundamental aims of the criminal justice system, the various wrongs and harms that that system would be inflicting on someone it convicts erroneously, the relative importance of furthering those aims and avoiding those wrongs and harms, and thereby come to a better understanding of the principles by which our judgements about culpability should be guided in conditions of uncertainty. I don't claim to have come to any such comprehensive understanding in this paper: the point that I have been at pains to establish is more that any such understanding will need to reflect on the importance of the needs, interests and values that underpin the criminal justice system and of the harms that the criminal justice system brings about when it gets it wrong. My argument for this claim chimes with the conclusion drawn by Alec Walen in his reflections on BARD (Walen 2015). We need to come to an evaluative conclusion – Walen recommends a retributive view – about the importance of increasing the number of convictions, given that this will further whichever ends we think that the criminal justice system ought to have, but given also

that any standard of proof will foreseeably result in false acquittals and false convictions. Only in grappling with these foundational questions, and by considering, as with the case of sexual attacks, how important criminal justice might be in the face of particular moral challenges we face at the moment, can we understand how best to frame the notion of reasonable doubt. And only by so framing that notion can we in the end decide how to approach the question of the fallibility foreseeably introduced into eyewitness testimony by the misinformation effect.

5. Conclusion

In this chapter I have looked at the way in which the criminal justice system should respond to the fact that the misinformation effect can distort eyewitness testimony. I contrasted two strategies, one that argues that it is sufficient to give juries guidance about the possibility of error, and another that argues that cases that rest on the validity of uncorroborated eyewitness evidence should be disallowed. I looked at Katherine Puddifoot's recent attempt to pursue the first strategy. However, I have argued that any such attempt will be lacking unless it confronts the foundational question of how important it is to do justice, and what matters when we get it wrong. Only then will we know how to attach credences to beliefs in cases in which, as the possibility of the misinformation effect shows, we do not have certainty but must always live with some level of probability that we have got it wrong.[2]

REFERENCES

[2] I am grateful to Zach Hoskins for helpful comments on a previous version of this paper, and to the participants in the 'Knowledge, Truth and the Criminal Trial' workshop at the University of Nottingham in June 2017.

Card, C. (1993). Rape as a Terrorist Institution. In: R. G. Frey and C. W. Morris, ed., *Violence, Terrorism and Justice*. Cambridge: Cambridge University Press, pp. 296-319.

DeRose, K. (1992). Contextualism and Knowledge attributions. *Philosophy and Phenomenological Research*, 52, pp. 913-929.

Devlin Report (1976). *Report to the Secretary of State for the Home Department of the Departmental Committee on Evidence of Identification in Criminal Cases*. London: HMSO

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12, pp. 361- 366

Fantl, J. and M. McGrath (2002). Evidence, Pragmatics and Justification. *Philosophical Review* 111, pp. 67-94

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

McMurtrie, J (2007). 'Incorporating Elizabeth Loftus's Research on Memory Into Reforms to Protect the Innocent. In: M. Garry & H. Hayne, eds., *Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom*. New York: Psychology Press. pp. 35-63

Puddifoot, K. (2018), 'Re-evaluating the Credibility of Eyewitness Testimony: the Misinformation Effect and the Overcritical Juror,' *Episteme* First View

Simon, W. H. (1993). The Ethics of Criminal Defense. *Michigan Law Review*, 91, pp. 1703-1728.

Stanley, J. (2007). *Knowledge and Practical Interests*. Oxford: Oxford University Press.

Walen, A. (2015). Proof Beyond a Reasonable Doubt: A Balanced Retributive Account. *Louisiana Law Review*, 76, pp. 355-446

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2007). Misinformation Effects and the Suggestibility of Eyewitness Memory. In: M. Garry & H. Hayne, eds., *Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom* (pp. 35-63). New York: Psychology Press.