

# Re-evaluating the Credibility of Eyewitness Testimony: The Misinformation Effect and the Overcritical Juror

## Abstract

Eyewitnesses are susceptible to recollecting that they experienced an event in a way that is consistent with false information provided to them after the event. The effect is commonly called the misinformation effect. Because jurors tend to find eyewitness testimony compelling and persuasive, it is argued that jurors are likely to give inappropriate credence to eyewitness testimony, judging it to be reliable when it is not. It is argued that jurors should be informed about psychological findings on the misinformation effect, to ensure that they lower the credence that they give to eyewitness testimony to reflect the unreliability of human memory that is demonstrated by the effect. Here I present a new argument, the overcritical juror argument, to support the conclusion that eyewitnesses are likely to make inappropriate credence assignments to eyewitness testimony. Whereas previous authors have argued that jurors will tend to give too much credence to eyewitness testimony, I identify circumstances in which jurors will give too little credence to some pieces of testimony. In my view jurors should be informed by psychological findings relating to the misinformation effect to ensure that they do not lower the credence that they give to eyewitness testimony when they should not.

## 1. Introduction

Testimony from eyewitnesses can be decisive in criminal trials.<sup>i</sup> However, over the last hundred years psychologists have gathered a vast body of evidence that is frequently taken to show that the testimony is unreliable (Howe and Knott 2015). The claim that eyewitness testimony is unreliable has been based in part on research showing that people are susceptible to what is commonly called the *misinformation effect*.<sup>ii</sup> In the misinformation effect, a person recollects that they experienced an event in a way that is consistent with false information provided to them after the event. It is widely accepted that people tend to find eyewitness testimony compelling and persuasive (e.g. Brigham and Bothwell 1983), but psychological results on the misinformation effect suggest that the memories that are the source of eyewitness testimony can be distorted, suggesting that people should approach the testimony with more scepticism than they are predisposed to. It has been argued that jurors and others working with the legal system should be introduced to psychological findings illustrating the misinformation effect to ensure that credence assignments to eyewitness testimony are appropriately lowered.

In recent work, however, Kourken Michaelian (2013) has argued that, in spite of the evidence widely discussed in psychology and legal theory, the belief forming processes leading to the production of eyewitness's memories and then testimony tend to produce true beliefs. Based on Michaelian's argument, it might appear that jurors act appropriately when they assign high credence to eyewitness testimony because eyewitness's memory beliefs will tend to be true. As arguments in favour of the introduction of psychological findings to court have been driven by the goal of ensuring that jurors reduce the credence that they assign to eyewitness testimony to reflect the *unreliability* demonstrated by the effect, Michaelian's argument seems to suggest that the psychological findings need not be introduced. If eyewitness's memories and testimonies about witnessed events are reliable, then *prima facie* there is no reason to introduce psychological findings.

In this paper I reject this result. I provide a new argument that supports the conclusion that jurors often assign inappropriate credence to eyewitness testimony. My argument highlights the importance of introducing psychological findings on the misinformation effect into the criminal trial to ensure that jurors give eyewitness testimony appropriate credence. However, my argument does not depend on establishing that eyewitness testimony is unreliable and that the credence given to this testimony needs to be lowered. Instead I argue that jurors are likely to give *too little credence* to some individual pieces of eyewitness testimony. While the existing literature on the misinformation effect emphasises how psychological findings can be used to encourage jurors to lower the credence that they give to eyewitness testimony, I emphasise how knowledge of the findings can prevent jurors from lowering the credence that they give to eyewitness testimony when they should not do so.

I argue that where evidence has been presented suggesting that an eyewitness has provided testimony that includes inaccurate details, jurors are likely to lower the credence that they assign to all or a substantial part of an eyewitness testimony, concluding that the eyewitness is either motivated to deceive or lacking a good supply of true beliefs about the event about which they are testifying. However, if the error is due to the misinformation effect then the eyewitness is likely to have a good supply of true beliefs that she is motivated to provide via her testimony. In these types of cases, the juror would be more likely to make a correct credence assignment, because she would be *less likely* to lower the credence that she gives to *all* or a *substantial part* of the eyewitness's testimony, if she were aware of the psychological findings on the misinformation effect. These psychological findings would show that people can make errors in their testimony (due to the misinformation effect) but nonetheless be trustworthy sources of much information about the case. And the same psychological studies could aid the juror in identifying whether, for any individual piece of information provided via testimony, errors are likely to have been made.

## **2. The misinformation effect and (un)reliable eyewitnesses**

There is now a vast body of psychological literature on the misinformation effect, showing how eyewitness testimony can become distorted as a result of information presented to the witness after they have witnessed an event. For example, Elizabeth Loftus and J.C. Palmer (1974) presented experimental participants with images of a car crash. Some participants were then asked how fast the cars involved in the crash were travelling when they "hit" each other while others were asked how fast the cars were travelling when they "smashed" into each other. Those who were asked the question using the word "smashed" were more likely to falsely report that there was broken glass in the image. In another experiment, participants were also shown an image of a car at a "stop" sign and then supplied with the misinformation that there was a "yield" sign (Loftus et al 1978). Participants provided with the misinformation were more likely than controls to claim that they recalled seeing a "yield" sign. In a more recent study, participants were shown a complex event, e.g. a girl having her wallet stolen by a man (Okado and Stark 2005). Then they were presented with misinformation about the event, e.g. that the man hurt the girl's arm when it was her neck that had been hurt. The misinformation was remembered as being a part of the original event 47% of the time. Hundreds of studies have now been undertaken demonstrating the robustness of this phenomenon (Howe and Knott 2015), which has

become labelled the *misinformation effect* because it occurs due to the influence of misinformation provided after an event

The misinformation involved in the effect can come from a variety of sources (Loftus 2005). Many psychological experiments, including those described above, illustrate how misinformation can be provided through suggestive interviewing by police officers or lawyers. For example, a police officer could refer to some detail (e.g. a “stop” sign) in her questioning, leading a witness to falsely remember that they saw the detail. However, a *social contamination effect* has also been established: it has also been found that people’s memories can be influenced by information provided by unfamiliar social peers (Gabbert et al 2004). In one experiment exemplifying this effect, participants were shown a video of a simulated car crash and then supplied with misinformation about the crash (ibid). Participants were significantly more likely to answer questions about the crash in a way consistent with the misinformation that they were supplied when the confederate, who appeared to be a social peer, provided it than when the information was provided in the form of a written narrative. The contamination effect is important when it comes to eyewitness testimony because witnesses often discuss with each other what they recall of a crime scene (ibid.). This provides an opportunity for them to contaminate each other’s memories of the details of the crime.

The misinformation effect leads to distorted memories. When these distorted memories are about the details of a criminal case, and expressed in eyewitness testimony, they lead the testimony to include inaccurate details. As a result, findings on the misinformation effect have “challenged prevailing views about the validity of memory and raised serious concerns about the reliability of eyewitness testimony” (Zaragoza et al 2007: 37). They have led the leading researcher on the misinformation effect, Elizabeth Loftus and her colleague to conclude that “[...] Eyewitness testimony is very powerful and convincing to jurors, even though it is not particularly reliable”(Laney and Loftus 2017). In a recent overview of research on memory distortion and eyewitness testimony, the impact of the research was summarized as follows:

For decades, psychologists and defense attorneys have maintained that eyewitness testimony can be notoriously unreliable, and courts, including the United States Supreme Court, have recognized this fact (Wise et al 2009).

In presenting psychological evidence in support of their view that eyewitness testimony is unreliable, these authors provide reason for jurors to lower the credence they give to eyewitness testimony to reflect the potential for eyewitness error.

In response to the findings on the misinformation effect, questions have been raised about the ability of jurors to determine whether eyewitness testimony is accurate and should be depended upon to formulate a correct verdict (e.g. Schmechel et al 2006). As jurors tend to find eyewitness testimony to be compelling and persuasive (e.g. Brigham and Bothwell 1983), but the testimony is subject to distortion due to the misinformation effect, jurors will tend to find evidence that can be distorted due to the misinformation effect to be compelling and persuasive. Under such conditions, jurors overestimate the adequacy of the testimony. Psychologists and legal theorists therefore argue that jurors should be informed about the psychological findings, so

that they are aware that eyewitness testimony can result from distorted memory, and the credence given to the testimony can be reduced accordingly.<sup>iii</sup>

In contrast, the philosopher Kourken Michaelian (2013) presents a challenge to the claim that eyewitness testimony is unreliable. Michaelian argues that the process of incorporating information acquired after an event via testimony is a reliable one, so the misinformation effect reliably produces true beliefs. Michaelian presents findings suggesting that the information that is provided after an event, influencing the recollection of the event, is often accurate. He cites a study undertaken by De Paulo and colleagues (1996) suggesting that in the majority of interactions people do not intentionally tell untruths, and that the vast majority of lies are about the liar, presented for psychological reasons. A further study cited by Michaelian suggests that people tend to tell the truth as a default, choosing to lie only when telling the truth would hinder their goals, for example, leading to some degree of social awkwardness, tension, or discomfort (Levine et al 2010, cited by Michaelian 2013).

What these studies suggest is that people tend to tell the truth when they have no motive to deceive. Moreover, where people are deceptive, it is usually in presenting information about themselves. This means that as long as the people sharing information about a crime to an eyewitness are knowledgeable about the subject matter that they are discussing, and are not talking about themselves, or aiming to achieve some goal that they cannot achieve through truth-telling, they are likely to present accurate information. If this information is integrated with other information deriving from memory, then the product is likely to be true beliefs about the crime that has been witnessed. The incorporation of information provided by others after experiencing a crime will therefore be a reliable belief forming process. Michaelian concludes that:

given the operation of the honesty bias, most [...] of the testimony received by an agent will be true; and if most of the testimony received by an agent is true, then [...] the unknowing incorporation of testimonial information into episodic memory is a reliable process for the formation of memory beliefs (2013, 2450).

Michaelian's argument challenges the idea that eyewitness testimony is unreliable, suggesting that the memory beliefs on which eyewitness testimony depends tend to accurately reflect the reality of the crime because they are influenced by the true testimony that others have provided to the eyewitness.<sup>iv</sup>

Michaelian's argument also provides reason to doubt that jurors assign inappropriate credence to eyewitness testimony. It might be argued that it is appropriate for jurors to take eyewitness testimony to be compelling and persuasive if the testimony tends to accurately reflect the reality of the crime that the testimony is about. If jurors assign appropriate credence to eyewitness testimony without being informed by psychological findings on the misinformation effect then there will be no benefit to introducing these findings into the courtroom. Therefore, Michaelian's argument appears to present a challenge to both the idea that jurors assign inappropriate credence to eyewitness testimony and the idea that jurors are more likely to assign appropriate credence to eyewitness testimony if they are informed by psychological findings on the misinformation effect.

What, then, should be thought about jurors' responses to eyewitness testimony? Do jurors assign inappropriate credence to eyewitness testimony? Should psychological findings on the misinformation effect be introduced into the courtroom to prevent this error? It might seem that the debate between psychologists like Loftus and Michaelian needs to be settled to answer these questions. However, in this paper I do not aim to settle this debate. Instead, I propose a new argument that shows that jurors *are* likely to assign inappropriate credence to some eyewitness testimony. The argument does not depend on existing arguments in support of the view that eyewitness testimony is unreliable, nor upon arguments claiming to establish that it is reliable. However, as we shall see in section 9, the argument does provide new support for the conclusion that eyewitness testimony is more reliable than has been appreciated by psychologists who argue that the credence assigned to eyewitness testimony should be lowered. The argument therefore occupies an interesting position relative to the existing debates: I agree with psychologists like Loftus that jurors assign inappropriate credence to eyewitness testimony *and* with Michaelian that eyewitness testimony is more reliable than it is often taken to be.

### **3. A new look at juror credence assignments**

The new argument I propose is the *overcritical juror argument*. According to the overcritical juror argument psychological findings on the misinformation effect should be utilised within the legal system *not* because this could ensure that the credence given to eyewitness testimony is lowered appropriately to reflect the potential for error due to the misinformation effect but instead because this could *prevent the credence* given to individual pieces of eyewitness testimony from *being lowered inappropriately*.

For the sake of the current discussion, a consequentialist approach is adopted when discussing what it is to lower credence appropriately. It is assumed that it is appropriate to lower the credence given to an individual piece of eyewitness testimony only if, in so doing, one increases the chance of a correct verdict being reached about the case in which the testimony is used. One inappropriately lowers the credence given to an individual piece of eyewitness testimony if, in so doing, one decreases the chance of a correct verdict being reached. My claim is that the psychological findings on the misinformation effect present very good reason to think that the credence given to individual pieces of eyewitness testimony is frequently lowered in a way that reduces the chance of a correct verdict being reached. On a more positive note, the same psychological findings can be utilised within the legal system to prevent errors of this type from occurring.

### **4. The cases**

The cases that are of interest are those in which an eyewitness appears to be discredited because of a single or small number of errors in their testimony. The following are examples. 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. In the specific cases of interest, jurors who are assessing the credibility of an eyewitness's testimony conclude one or both of two things: (i) the eyewitness

intends to deceive by providing false details; (ii) the eyewitness does not have a good supply of true memories about the event about which they are testifying. This section establishes that these types of reactions occur.<sup>v</sup>

Let us first consider how jurors might come to conclude that an eyewitness has an intention to deceive on the basis of evidence of errors in their testimony. One important role of jurors is to detect whether an eyewitness intends to deceive (see, e.g. Seniuk 2013), but people, including those working within the legal system, are generally no better than chance at discerning whether someone has honest or dishonest motives in providing information (Bond and DePaulo 2006; Ekman and O'Sullivan 1991). One clue that is often taken to indicate that a person who is testifying is being dishonest is inconsistencies or errors in their testimony (ten Brinke and Porter 2013). And people who adopt an initial position of mistrust towards an individual are especially likely to view errors as evidence of an intention to deceive (Meissner and Kassin 2004). The nature of criminal cases is that stakes are high and there can be numerous actors with competing goals, so jurors will often adopt a starting point of suspicion, suspecting any eyewitness of having a motivation to deceive. When these factors are combined, jurors are likely to take evidence of errors to indicate that a person has an intention to deceive. This is especially likely where the erroneous evidence is perceived as providing support for a verdict that the eyewitness is expected to favour. A juror could then lower the credence given to all or many substantive parts of the eyewitness's testimony, that is, those parts of the testimony that indicate something significant about the guilt or innocence of certain suspects, on the basis that they think that the eyewitness intended to deceive when providing the details.

Now let us consider how jurors might come to conclude on the basis of evidence of a single or small number of errors that an eyewitness does not have a good supply of true memories about the event about which they are testifying. Support for this claim comes from a series of psychological studies on the effect of eyewitness error. The studies show that where an eyewitness provides details that are refuted the overall credibility of the eyewitness testimony is reduced. For instance, where participants in an experiment were given two sets of contradictory evidence from fictional eyewitnesses about a fictional car accident, and one set included trivial details that were refuted, the testimony of the eyewitness who provided those details was subsequently given less credibility (Borckardt, Sprohge and Nash 2003). A similar effect occurred when experimental participants were shown testimony containing inconsistencies; they found the eyewitness to be less effective and were unlikely to convict on the basis of the testimony (Berman and Cutler 1996; Hatvany and Strack 1980), even when the inconsistencies related to trivial details (Berman, Narby and Cutler 1995). It is important that testimony was discredited, and not taken to provide the support for a particular verdict that it would otherwise seem to be provide, when the errors that were demonstrated related to *trivial* information. The truth or falsity of the trivial details would not establish guilt or innocence. Nonetheless, those fulfilling the juror role were less likely to convict when the trivial details were shown to contain errors or be inconsistent with other evidence. This shows that errors in some details within an eyewitness's testimony, i.e. the trivial details, are taken to undermine the credibility of other details, i.e. those that indicate guilt or innocence. Eyewitnesses are assumed, based on a small number of trivial errors, to lack a generally good supply of true memories. These experimental findings thus provide good reason for thinking

that that jurors are susceptible to concluding that an eyewitness lacks a good supply of true beliefs about a case based on evidence of a single or small number of errors.

### **5. Eyewitness error and the intention to deceive**

In the next three sections I aim to show that the juror who responds to evidence of error in eyewitness testimony by concluding (i) or (ii) will be susceptible to assigning inappropriately low credence to various of the eyewitness's beliefs when the error is due to the misinformation effect. This section focuses on how the conclusion that evidence of error shows that the eyewitness has an intention to deceive will be false when their error is due to the misinformation effect. To see this point, it is useful to reflect upon the explanation of the effect given in the cognitive sciences, and how it implies that the effect occurs without being caused by a motivation to deceive.

The misinformation effect is argued by cognitive scientists to be the result of the ordinary operation of human cognitive mechanisms. The effect has been explained by appeal to the constructive nature of memory (see e.g. Loftus 2005; Michaelian 2013; Schacter et al. 2011). According to *constructivism*, traces of information rather than discreet records of events are stored to memory and the traces are combined at the point of retrieval to construct a representation of a past event. This phenomenon is taken to explain the misinformation effect in the following way: Traces of information about one's personal experience of an event can become combined with traces of information of being provided with testimony or information about being suggestively questioned so that one falsely recalls events in a way that is consistent with information provided after the event (ibid.). Where the information that is provided after the event by others is *misinformation* the result is the misinformation effect: distorted memories are formed that are consistent with the misinformation rather than genuinely reflecting the event as experienced.

As such, the effect does not occur due to any intention to deceive on the part of the believer. Some memory distortions seem to have motivational components, for example, memory systems can filter incoming information in ways that are self-enhancing, supporting a positive view of the person doing the remembering (Wilson and Ross 2003; Sutton 2009). However, even in these cases, the motivational component of the distortion is not taken to involve any consciously chosen intention to deceive. Moreover, the misinformation effect in particular is explained at the level of cognitive mechanism without reference to the motivation of the person who possesses the cognitive mechanism (see, e.g. Schacter et al 2011). This means that even an eyewitness testifier who is strongly driven by the desire to provide true testimony, and who has no conflicting desires, can be susceptible to the effect.

As errors in the details provided within eyewitness testimony can be the result of the misinformation effect, and the misinformation effect is not caused by an intention to deceive, errors in the details of testimony can occur in the absence of an intention to deceive. Jurors who assess the credibility of an eyewitness who provides erroneous testimony due to the misinformation effect and conclude that they intend to deceive are therefore likely to inappropriately lower the credence that they give to much of the eyewitness's testimony.

### **6. Eyewitness error and a good supply of true memories**

Let us now consider the second conclusion that could be drawn on the basis of evidence of error in an eyewitness account: that the eyewitness does not have a good supply of true beliefs about the event about which they are testifying. This section shows that it is consistent with evidence of error in an eyewitness account that occurs due to misinformation effect that the eyewitness has a good supply of true beliefs about the event that they could provide in their testimony.

It is, first of all, consistent with evidence of error in an eyewitness's account that she is intellectually virtuous (Montmarquet 1987, 1993; Zagzebski 1996).<sup>vii</sup> In section 5 it was shown that the best existing explanation of the effect suggests that an eyewitness can be properly motivated to provide true testimony but nonetheless make errors due to the effect. This means that the eyewitness can be conscientious in the way that she attempts to recall information from her memory, she can be open-minded about what happened, she can be driven by the aim to pay attention to detail and display many other intellectual virtues (Montmarquet 1987, 1993; Zagzebski 1996), and so on, but nonetheless make errors in her recall of an event due to the misinformation effect. As long as there is good reason to think that a person who displays intellectual virtues of this kind is likely to have many true beliefs about any particular event, it is consistent with errors due to the misinformation effect that the eyewitness can have a good supply of true beliefs about an event as a result of being intellectually virtuous.

Second of all, a person is likely to have a good supply of true beliefs about the *gist* of an event that they have experienced even if they lack true beliefs about the details of that event. For example, an eyewitness might falsely recall that a person involved in an accident passed a "stop" sign rather than a "yield" sign but nonetheless accurately recall that one person travelling at a dangerous speed caused the accident. What reason is there for thinking this? Psychological studies suggest that human memory systems form two sorts of representations of events: *gist representations* and *verbatim representations* (Brainerd & Reyna 2002). This fact is taken to explain the phenomenon found in the Deese-Roediger-McDermott series of experiments (developed by Deese 1959, and revised by Roediger & McDermott 1995). In these experiments participants are shown a list of words (e.g. *baker, butter, filling, brown, dough, grain, flour, knife, wheat, old*) and asked to recall them. They systematically and predictably claim to have studied words that are not on the list but are semantically related to words on the list (e.g. *bread*). This phenomenon is explained in the following way: the participants form a gist representation of the items in the list (e.g. *bread-related items*) as well as a verbatim representation (containing each individual item). When they come to recall the items they utilise the gist-representation, filling out the details in the list in a way that fits the common theme of the list. They make the error of claiming that they studied unstudied words because they include items that fit the theme captured by the gist representation but were not on the original list. It is thought that the gist-representation remains while the verbatim representation fades (Brainerd & Reyna 2002), allowing the rememberer to retain key information even as their grasp of the details fade. The dissociation between the two forms of representation means that an eyewitness might lack an accurate representation of some of the details of an event but nonetheless be a good source of information about the gist of the event.



Psychological studies also suggest that the eyewitness will have a good supply about many details of an event other than those about which she has provided erroneous testimony. The studies suggest that at least one of a number of “eliciting” or triggering conditions need to be in place for the misinformation effect to occur (Loftus 2005), and, for any event, these conditions might be in place for some details about the event but not others. Most obviously, for the misinformation effect to occur, misinformation about the details of an event must be made available to the person who misremembers. For example, an eyewitness must be exposed to suggestive questioning or provided misinformation from another eyewitness about the specific detail. A person can be provided misinformation about some particular detail of an event (e.g. the type of sign that a car passed through) but not some other detail (e.g. the colour of the car). She can consequently be subject to the misinformation effect with respect to her memories of some details of the event but in a position to provide accurate details about others.

Other factors that determine whether a person is susceptible to the misinformation effect include: (i) the time that passes between them being provided with misleading information and them being asked about the event (Loftus et al 1978); (ii) whether or not they think that they were in a good position to form an accurate memory (e.g. whether or not they think they were drunk when the incident occurred) (Assefi and Garry 2002); (iii) the social status of the person who provides the misinformation (Dodd and Bradshaw 1980; Smith and Ellsworth 1987; Underwood and Pezdek, 1998); (iv) whether or not the accent of the person who provides the misinformation conveys that they are powerful and socially attractive (Vornik et al 2003); (v) whether or not there is a strong discrepancy between the person’s memories of the event and the information provided about the event (Tousignant et al 1986); (vi) whether the person pays attention at the point of encoding the information and the point of retrieval (Zaragoza and Lane 1998); (vii) whether or not the person is repeatedly exposed to the misleading information (Zaragoza and Mitchell 1996; Mitchell and Zaragoza, 1996).

For any event, some of these triggering conditions might be present for some details of an event but not others. For example, one might undergo repeated questioning about whether a person accused of dangerous driving passed through a “stop” sign without stopping, while the colours of the cars involved in the accident are never mentioned to you. Or you might confidently remember the make and model of the car, because you are a car mechanic so think that you were in a good position to form an accurate memory of this detail, and vehemently reject the suggestion that the car was another make and model, thereby avoiding the misinformation effect. But you might simultaneously be susceptible to misremembering that there was a “stop” sign because you did not attend to the street signs at the scene of the crime. In cases of this sort, psychological studies suggest that some details of an event are likely to be misremembered due to the misinformation effect while other details are remembered accurately.

Furthermore, there is reason to think that people will often avoid the misinformation effect even when they could be susceptible (Michaelian 2011). Source-monitoring errors are a leading explanation of the misinformation effect (for reviews see e.g. Belli and Loftus 1994; Lindsay 1994; Mitchell and Johnson, 2000). The misinformation effect involves people misidentifying memories that have their source

in testimony as memories that have their source in experience. But cognitive scientists argue that source-monitoring errors are exceptional because traces of information stored in memory bear the mark of their origins (Johnson 1988). For example, memory representations of events that have been perceived are usually experienced as more vivid than memory representations of events that are dreamed (Mitchell and Johnson 2000). 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 (ibid). The experiential differences between the representations ensure that people are often good at identifying their source through this comparison (Michaelian 2011). The process is not infallible, of course: the fallibility of the process explains how the misinformation effect occurs, but it is thought that the process is reliable because the marks of origin reliably indicate the source of memories (ibid.). Work on source-monitoring thus suggests that people often avoid the misinformation effect because they can properly identify the source of some misinformation that they have been supplied. Consequently, evidence that a person has made an error that is the result of the misinformation effect does not mean that she is likely to make numerous other errors.<sup>viii</sup>

In sum, then, findings from the cognitive sciences strongly suggest that a person can display errors in her eyewitness testimony due to the misinformation effect while nonetheless having a good supply of true beliefs about the event about which she is testifying. She might (a) have an excellent intellectual character; (b) be able to accurately remember the gist of what happened in the event; and (c) be able to accurately remember many details. This means that jurors who conclude on the basis of evidence of errors in an eyewitness's testimony about an event that she lacks a good supply of true beliefs relating to the event are likely to inappropriately lower the credence they give to some details in the testimony.

## **7. Errors and the Ordinary Functioning of Good Eyewitnesses**

The argument presented so far provides reason for accepting that it is consistent with an eyewitness providing testimony that includes false details that she has a good supply of true beliefs that she is motivated to provide in the witness stand. This section shows that, 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.

To substantiate this point, I first outline various descriptions of the functions of the cognitive mechanisms that produce the misinformation effect provided in the cognitive sciences and philosophy of memory (in section 7.1). I show that all of the functions bring distinctive epistemic benefits (section 7.2). I argue that the mechanisms are *epistemically innocent*: although they bring epistemic costs they also bring substantial epistemic benefits that would otherwise be missed (section 7.2). Then I show that the benefits facilitate humans, who each possess the mechanisms, being good eyewitnesses (section 7.3).

*Epistemic benefits* are defined for the purposes of the current discussion as features of a cognitive system that facilitate the achievement of epistemic goals “including acquiring new true beliefs; retaining and using relevant information; increasing the coherence of a set of beliefs; and gaining understanding” (Puddifoot and Bortolotti 2018). There has been a great deal of discussion within the cognitive sciences about how memory distortions, such as the misinformation effect, are the result of the ordinary operation of cognitive mechanisms that are *adaptive* (see, e.g. Schacter et al. 2007; Sutton 2009; Schacter 2011). However, the current discussion focuses on how the cognitive mechanisms responsible for the memory distortion bring specifically *epistemic* benefits.

### **7.1. Functions of constructive memory systems**

As discussed in section 5, the misinformation effect is explained by leading cognitive scientists as being a result of the *constructive nature of human memory systems* (see, e.g. Loftus 2005; Michaelian 2013; Schacter et al. 2011). The *archival view of memory*, according to which memory functions like a storehouse, storing complete and discreet records of events, is now generally rejected and a variety of versions of *constructivism* have come to dominate the cognitive science and philosophy of memory (Robins 2016, for defences of constructivism see e.g. Bartlett 1932; Neisser 1967; Suddendorf and Corballis 1997, 2007; Loftus 2005; Schacter and Addis 2007, Schacter et al 2007, Schacter et al 2011; Shanton and Goldman 2010; Michaelian 2011, 2013, 2016a, 2016b; Klein 2013; De Brigard 2014).

Constructivists argue that the cognitive mechanisms responsible for memory support functions other than accurately representing the past (Suddendorf and Corballis 2007; Schacter and Addis 2007; De Brigard 2014; Michaelian 2016a,b). According to constructivism, the best way to explain the frequency of the systematic and predictable memory errors studied in the cognitive sciences—including the misinformation effect—is to suppose that the production of accurate representations of the past is not the sole, or even the main, function of the cognitive mechanisms that support remembering (Sutton 1998; Michaelian 2012; De Brigard 2014).<sup>ix</sup>

Constructivists appeal to evidence suggesting that people with pathologies like Alzheimer’s are less prone to some common memory errors that are the result of the process of construction (Schacter et al 1996; Schacter et al 1997; Melo et al. 1999). This evidence is taken to show that failing to accurately represent the past is a part of the proper functioning of the cognitive mechanisms that underpin memory—proper functioning that is disturbed in cases of the pathology involved with Alzheimer’s. If inaccurately representing the past is a part of the proper functioning of the cognitive mechanisms responsible for memory then accurately remembering the past cannot be the sole function of those mechanisms. Furthermore, constructivists appeal to evidence suggesting that susceptibility to common memory errors correlates with the ability to successfully complete some cognitive tasks, including those involving convergent thinking (Howe et al 2010; Dewhurst et al 2011). Each of these findings are taken to support the conclusion that the cognitive mechanisms that produce memories, including the distorted memories produced as a result of the misinformation effect, subserve a broader cognitive function, encapsulating but not limited to remembering past events.

There is some disagreement among constructivists about the precise functions that are subserved by the cognitive mechanisms that produce memories. What follows is a survey of the constructivist positions.

On the mental time travel view, the goal of the cognitive mechanisms is mental time travel: *the projection of the self into the past or future* (Tulving 1985; Suddendorf and Corballis 1997, 2007).

[...]Episodic memory implies a mental reconstruction of some earlier event, including at least some of the particularities of that event, such as the principal characters involved, the actions that took place, the setting, and the emotional reactions. Metaphorically speaking, it might be regarded as the result of a mental journey into the past. This idea is readily extended to the future. Based on previous experiences, we can imagine specific events in the future, including the sorts of particularities that have characterized events in the past (Suddendorf and Corballis 2007: 301).

On the mental time travel view, remembering involves mental time travel into the past, but it is unlikely that the cognitive mechanisms evolved to support time travel in this direction. The adaptive advantage of mental time travel comes from its facilitation of time travel into the future: i.e. the ability to experience how future events are going to be (Suddendorf and Corballis 2007). By accurately predicting the future, mental time travel *supports good decision-making, which increases fitness*.

The following evidence is used to support the mental time travel model: (i) Evidence that increased distance of an event from the present reduces the phenomenological richness of the experience of the event regardless of whether the event is in the past or future (D'Argembeau & Van der Linden 2004). (ii) Evidence that patients with amnesia who are unable to answer questions about yesterday are also unable to answer questions about tomorrow (Klein et al. 2002; Tulving 1985). (iii) Evidence that patients with depression who struggle to retrieve information about specific events in their past also struggle to imagine future episodes (Williams et al. 1996). (iv) Evidence that children gain the ability to answer questions about yesterday and tomorrow at the same age: around 4 years old (Busby & Suddendorf 2005). (v) Brain-imaging studies suggesting that the same "core network" of neural regions are utilised in remembering the past and predicting the future (Okuda et al. 2003). Each of these findings suggests that the ability to project oneself into the past and future are supported by the same cognitive mechanisms. A plausible evolutionary explanation of the existence of the cognitive mechanisms is that their primary function is to facilitate the *projection into the future required for good quality decision-making*.

Daniel Schacter and Donna Addis (2007, see also Schacter et al. 2007; Schacter et al. 2011) defend a similar view of the functions of the cognitive mechanisms supporting memory to that of the mental time travel model, but focus more closely on the operation of constructive processes. According to their *constructive episodic simulation hypothesis*, the primary function of human memory systems is *the flexible simulation of future events*.<sup>x</sup> The constructive nature of human memory facilitates accurate simulation of future events because the future never precisely resembles the past. Constructive cognitive mechanisms allow information from various different events to be abstracted and combined to simulate novel future events. Although the

process of construction facilitates accurate simulations of the future, it also explains errors such as the misinformation effect because the process of construction can involve “miscombining elements” (Schacter et al. 2011), including from memory and testimony.

The idea that the function of memory is to simulate the future has received uptake from philosophers of memory defending simulationist approaches to memory (Shanton and Goldman 2010; Michaelian 2011, 2016a,b). On the *simulationist view*, remembering is “generating a more or less probable representation of a target event” (Michaelian 2016b, 6). Successful remembering can be distinguished from unsuccessful remembering, such as confabulation, on the basis that successful remembering is produced by memory systems operating in ways that tend to produce true simulations whereas confabulation occurs through memory systems operating in ways that tend to produce false simulations (ibid).

In defending his *episodic hypothetical thinking model*, Felipe De Brigard (2014) argues that the cognitive mechanisms that produce memories operate to meet a broader goal than that specified within the mental time travel and constructive episodic simulation hypothesis views: the *formation of simulations of what might happen or might have already happened to us*.<sup>xi</sup> De Brigard argues that his view is consistent with a vast body of evidence. It is consistent with the evidence suggesting that the same cognitive mechanisms underpin projection into the past and future, which also supports the mental time travel view and the constructive episodic simulation hypothesis. If the main function of the cognitive mechanisms is to develop representations of events that might happen or might have already happened to us then these representations could be used to project oneself into the past or future. The *episodic hypothetical thinking model* is also consistent with evidence that memory retrieval is sensitive to the probability of an event happening in a particular context. For example, the rate at which participants correctly or incorrectly recalled that an item was found in a context was predicted by their judgements of the likelihood that the item would be found in the context (Hemmer and Steyvens 2009). On this view, the way that memories are constructed reflects the probability that the event occurred (in any particular way): constructions are plausible representations of ways that things could be.

De Brigard’s view is distinct from the mental time travel view and constructive episodic simulation hypothesis because it suggests that the main functions of the cognitive mechanisms that produce memories include facilitating thinking about *counterfactuals*, i.e. what could have happened but did not. Support for this more inclusive view of the functions of memory systems comes from the following sources: (i) Evidence that children begin counterfactual thinking at the same time memories begin consolidating in long-term memory and when episodic future planning begins (German and Nichols 2003). (ii) Evidence of patients with deficits in both episodic memory and counterfactual reasoning, for example, evidence that people with amnesia provide impoverished descriptions of plausible imagined events (Rosenbaum et al 2009; Hassabis et al 2007). (iii) Evidence suggesting that there are similarities in the phenomenology and amount of details given by participants during thinking about episodes in the past and future and during counterfactual thinking (De Brigard and Giovanello 2012). (iv) Evidence that the same “core network” of brain regions are implicated in counterfactual thinking and recollection of autobiographical information

(Addis et al 2009; van Hoeck 2012). Each of these pieces of evidence is consistent with the cognitive mechanisms underpinning memory having as their main function the formation of simulations about what might happen or might have happened to us, including simulations of counterfactual events. On this view, the misinformation effect would occur due to the binding together of information from experience and testimony to formulate a plausible hypothesis about what might have happened in the past.<sup>xii</sup>

In sum, then, recent research from the cognitive sciences and the philosophy of memory suggests that memory errors such as those present in the misinformation effect are the result of the ordinary operation of cognitive mechanisms that support functions other than the representation of past events: (i) *projection into the future required for good quality decision-making*, (ii) *flexible construction of simulations of the future*, and/or (iii) *hypothetical thinking about what might happen or might have happened (including counterfactual thinking)*.

## **7.2. Epistemic Benefits and Epistemic Innocence**

Each of the functions ascribed to the cognitive mechanisms underpinning memory (and outlined in section 6.1) brings substantial epistemic benefits. Projection of oneself into the future facilitates numerous epistemic goals. It produces beliefs about the future. When successful, the projection produces *true* beliefs about the future. When it draws upon past experiences, projection into the future produces beliefs about the future that cohere with beliefs about the past. It allows one to navigate one's environment by setting expectations about features likely to be found in an environment. It can therefore facilitate the acquisition of knowledge and understanding downstream. For example, a scientist designing the methodology for an experiment might project herself into the future experimental setting. In the process, she might form true beliefs about the future. She might form beliefs about the experiment that fit with her beliefs about past experiments. As a result, she might have correct expectations about what measurements to make and data to gather. She might consequently obtain scientific knowledge and understanding as a result of successful mental time travel.

The construction of simulations of future events from traces of information about various different past events brings an extra epistemic benefit, in addition to the general benefits of mental time travel. It facilitates the flexibility of thought and creativity required to predict *novel* events, which are unlike any single event that has previously been experienced. Once these events are predicted, the agent will be in a good position to gain knowledge and understanding about the events and their consequences.

Engagement in hypothetical thinking about what might happen and what might have happened also facilitates the achievement of epistemic goals. Episodic hypothetical thinking facilitates the imagination of events that have not yet been experienced, providing a source of creativity, which allows people to determine what is likely to happen. It is therefore a source of knowledge and understanding. It focuses the attention of an agent on likely outcomes. When interpreting the past, making predictions about the future, or considering what might have been but was not, there are countless possibilities to countenance. Therefore, the task of selecting likely outcomes could be extremely computationally complex, intractable even. Episodic

hypothetical thinking formulates representations of only a subset of these possibilities: those that could plausibly have occurred. It therefore simplifies what could otherwise be an intractable computational task so that correct judgements are more likely to be made about what could have happened and what could happen in the future.

Due to the epistemic benefits associated with the ordinary functioning of the cognitive mechanisms that produce false memory beliefs like those produced by the misinformation effect, Katherine Puddifoot and Lisa Bortolotti (2018) have argued that the mechanisms are *epistemically innocent*. A cognitive mechanism is epistemically innocent if it meets the following conditions:

- (*Epistemic Cost*) Cognitive mechanism produces epistemically costly cognitive states.
- (*Epistemic Benefit*) Cognitive mechanism has some *significant* epistemic benefits for an agent.
- (*No Alternatives*) There is no available alternative cognitive mechanism that would enable the agent to avoid the epistemic costs while conferring the same epistemic benefits.

The memory mechanisms that produce the misinformation effect are epistemically costly because they lead to the production of false memory beliefs, for example, memory beliefs that incorporate false information provided via suggestive police questioning. Furthermore, when people have false memory beliefs, or even true memory beliefs that misrepresent the source of the beliefs (as in the misinformation effect), they lack self-knowledge. However, as has been demonstrated so far in section 7, on the best existing theories in the cognitive science and philosophy of memory the same cognitive mechanisms also bring significant epistemic benefits. The mechanisms facilitate the successful prediction of the future, allow the flexible recombination of information drawn from various past experiences, and/or allow hypothetical thinking about episodes that are likely to occur. Each of these functions brings substantial epistemic gains. Meanwhile, there are no alternative human cognitive mechanisms that operate like storehouses, storing complete and accurate records of past events, and therefore invulnerable to errors such as those involved in the misinformation effect. The cognitive mechanisms that produce the misinformation effect are therefore epistemically innocent.

The epistemic innocence of the cognitive mechanisms that produce memory errors, including through the misinformation effect, is important because it is intuitive that evidence that a mechanism produces false memory beliefs and is susceptible to falsely representing past experiences is only negative from an epistemic perspective (Puddifoot and Bortolotti 2018). People who display memory errors, including the misinformation effect, can be taken to be poor epistemic agents, and poor sources of information. However, memory errors such as those produced by the misinformation effect should not be taken to indicate that a person is a poor epistemic agent. The errors should not be viewed wholly badly from an epistemic perspective. This is because the memory errors can be an indication of the ordinary operation of cognitive mechanisms that bring substantial epistemic benefits, facilitating the achievement of important epistemic goals.

### 7.3. Epistemic Innocence and Eyewitness Error

The ideas of section 7.2 can now be applied to the case of eyewitnesses in criminal trials who make errors as a result of the misinformation effect. When they make the errors, there are subsequent epistemic costs: they misrepresent a detail of the crime that they have witnessed and lack self-knowledge about some of what they have experienced. However, the error is due to the ordinary operation of constructive cognitive mechanisms that are epistemically innocent: although the cognitive mechanisms produce errors, they bring significant epistemic benefits that would otherwise be missed. Moreover, importantly, the epistemic benefits make human beings able to be good eyewitnesses. Therefore, an eyewitness can display errors in their testimony due to the ordinary functioning of cognitive mechanisms that make them otherwise a good eyewitness, with a good supply of true beliefs to provide via their testimony. To illustrate this point, we will now consider each of the functions identified in sections 7.1 with the constructive cognitive mechanisms underpinning memory and how the epistemic benefits that they bring increase the chance of people being good eyewitnesses.

Being able to project oneself into the future is crucial to giving requisite attention to the features of a crime that occurs in one's presence, and to identifying what information about a crime will be relevant to future investigations. Imagine that there is a collision between two cars after one car passed over some traffic lights into a dangerous position at an intersection. What would make a person a good eyewitness in this case? She would notice that the car was passing into a dangerous position at the intersection. Then she would attend closely to the situation. She would be attentive to the colour of light that the car passed through, and the speed at which the car travelled, being able to predict that these are salient features of the situation. Once the accident occurred, she would be able to predict which features of the event would be important to future police investigations (colour of light, speed of travel, etc.). She might immediately mention these to the police to reduce the chance that she misremembers the details. She would engage in these activities, which would increase the chance of her being a good eyewitness to the case, by increasing the chance that she has a good supply of true beliefs about the case, because she was able to successfully predict the future.

The epistemic benefits of being able to project oneself into the future would here manifest as having *true beliefs* about what is likely to happen, what it is important to attend to, and what information should be passed on to the police. The downstream epistemic benefits of projection into the future would include the *knowledge* about the crime gained due to the proper attention paid to salient features of the situation.

Flexibly recombining pieces of information about various events facilitates being a good eyewitness by allowing people to draw on their own experiences of numerous different experiences to predict features of what will often be a novel event: the experience of witnessing a crime. Eyewitnesses will commonly not have had previous experiences of witnessing crimes, or, if they have had experiences of this type, the crimes and their experiences of witnessing those crimes will differ. They will therefore need to think flexibly, combining information about various different experiences that they have had that are similar in some respect to the current experience to make a prediction about what is likely to occur, what information is likely to be useful to criminal investigations, and so on.



Let us now consider the epistemic benefits of the final function associated with the cognitive mechanisms that produce memory beliefs: episodic hypothetical thinking. We found in section 7.2 that episodic hypothetical thinking brings the epistemic benefits of facilitating creative thinking about events that have not occurred and of reducing the number of possible outcomes an agent has to consider when thinking about what might happen by eliminating outcomes that are implausible. Creative thinking is crucial to imagining what is likely to happen in a crime, and what is likely to be relevant in a criminal trial. The tendency to selectively imagine only events that are likely to happen will increase the likelihood that an eyewitness has correct beliefs about what will happen, attends to appropriate information, identifies the information that will be relevant to a future inquiry, and so on. If an eyewitness considers only plausible outcomes of the crime and criminal process then considering less plausible outcomes will not distract her.

If existing theories within the cognitive sciences and philosophy of memory are correct, then, the cognitive mechanisms that produce memory errors through the misinformation effect bring substantial epistemic benefits other than representing the past *and* these benefits facilitate humans being good eyewitnesses. Human memory systems support mental time travel, flexible simulation of future events, and/or hypothetical thinking about what might happen or have happened. Each of these functions is epistemically beneficial in a way that allows human beings to be good eyewitnesses. This means that when a person displays errors due to the misinformation effect, these errors are the result of the ordinary operation of cognitive mechanisms that facilitate people being good eyewitnesses. Errors that are likely to be taken to show that an eyewitness lacks a good supply of true beliefs about the case about which they are testifying are actually explained by the presence of cognitive mechanisms that facilitate having a good supply of true beliefs about the case.

It is worth noting that the fact that someone has constructive cognitive mechanisms that facilitate mental time travel, flexible simulation of future events, or hypothetical thinking about what might happen or have happened does not mean that they will be a better eyewitness than other people. If two competing eyewitness testimonies are given, and one testifier is subject to the misinformation effect, then this does *not* mean that this testifier is more likely than the other to provide have a good supply of true beliefs about the case. This is because it is not the case that some people have constructive memory systems that bring the epistemic benefits identified in sections 7.2 and the current section and others do not—they will each have constructive memory systems, with the epistemic benefits they bring. On the other hand, because all eyewitnesses have constructive memory systems, they are all susceptible to the misinformation effect. Therefore, evidence that one eyewitness has been susceptible to the effect does not provide good reason to think that they are overall less reliable than another eyewitness.

There is substantial import to the observation that errors due to the misinformation effect are the result of the ordinary operation of human cognitive mechanisms that are epistemically innocent in ways that facilitate being a good eyewitness. It reveals the extent of the error that can be made by jurors when they infer from evidence of eyewitness error that is due to the misinformation effect that an eyewitness lacks a good supply of true beliefs. Not only do jurors err because it is consistent with the

eyewitness error that the eyewitness has a good supply of true beliefs, they err because the eyewitness errors are due to the operation of cognitive mechanisms that will, on very many occasions, ensure that a person has a good supply of true beliefs to deliver through their testimony.

### **8. Eyewitness error and poor credence assignments**

The discussion in sections 3-7 provides reason for thinking that jurors are likely to respond to evidence of erroneous details in an eyewitness's testimony by inappropriately lowering the credence given to the testimony. Where there is evidence that an eyewitness has made an error in her testimony, jurors are likely to conclude that she is either driven by the desire to deceive or that she does not have a good supply of true beliefs relating to the events about which she is testifying. A juror who responds in either of these ways, forming a belief about the trustworthiness of an eyewitness based on the evidence of error, and lowering the credence they give to all of the eyewitness's testimony, would be likely to lower the credence she gives to much of the testimony in a way that decreases the chance of a correct verdict being made. This is because, as has been shown in sections 4-7, it is consistent with the eyewitness making errors in some of her testimony that she can be an excellent source of information about other details relating to the case.

One way to respond to this conclusion is to argue that jurors should be discouraged from updating the credence that they assign to eyewitness testimony based on evidence of errors in their testimony. Or one might discourage jurors from responding to specific evidence of eyewitness error by forming the belief that the eyewitness is wholly untrustworthy. However, each of these responses risks reducing the chance of a correct verdict being made by allowing the testimony of witnesses who *are* motivated to deceive or *do* lack a good supply of true beliefs about a case to determine verdicts.

What is the solution? To change the supply of evidence that is available to the jurors, so that it includes information about the misinformation effect.<sup>xiii</sup><sup>xiv</sup><sup>xv</sup> By changing the epistemic environment of the jurors in this way, it will be possible to increase the chance they give testimony appropriate credence, by reducing the chance that they inappropriately lower the credence given to eyewitness testimony when an eyewitness makes an error due to the misinformation effect. As mentioned in section 2, the proposal that jurors should be informed by psychological findings relating to the misinformation effect is not new. 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.<sup>xvi</sup> 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. The achievement of the latter goal does not only require jurors to be aware that the misinformation effect occurs. It requires jurors to be aware about how representations of the gist of an event can be dissociated from representations of the details; about how a person can be a good source of information about some details relating to an event but not others; and about how memory errors due to the misinformation effect are indicative of the ordinary operation of cognitive mechanisms that facilitate being a good eyewitness. Moreover, it requires knowledge about the specific conditions under which people's memories are susceptible to distortion due to the misinformation effect.

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.<sup>1xvii</sup> 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. If it is concluded that the eyewitness is not likely to have been similarly questioned about other details that she provides then there is reason to think that her testimony about these details is accurate because it is less likely to be subject to the misinformation effect. The juror might have otherwise lowered the credence that she gave to the latter details, from, say, 0.9 to 0.1, due to evidence of errors in the eyewitness testimony. But once she is aware of the psychological findings on the misinformation effect, including about the conditions under which the effect is likely to operate, the juror might not lower the credence she gives to the testimony at all, leaving it at 0.9, if she remains confident that the conditions under which the misinformation effect is likely to operate are not present. She would thereby avoid lowering the credence given to the testimony inappropriately.

In sum, then, where jurors are predisposed to inappropriately lower the credence given to eyewitness testimony in response to evidence of error, if they are exposed to psychological findings on the misinformation effect, this predisposition can be changed. The juror who is equipped with information about the misinformation effect can prevent herself from making inappropriately low credibility assignments. There is consequently good reason for jurors and other participants in the legal system to be exposed to the psychological findings other than the reason often provided in the existing psychological literature.

## **9. Fallibility and Unreliability**

The current paper puts pressure on the claim commonly made in the existing literature that evidence of memory errors like those involved in the misinformation effect establishes that episodic memory systems are unreliable. The psychological findings discussed in section 6 and 7 demonstrate that episodic memory systems (i.e. those cognitive mechanisms responsible for memories about individual incidents) can produce a good supply of true beliefs even though they produce distorted memories. If the reliability of a cognitive system is determined by its capacity to produce a high ratio of true compared to false beliefs (Goldman 1979), then it is consistent with evidence of error like that caused by the misinformation effect that episodic memory systems are reliable. The memory systems could provide enough true beliefs about the gist of events and about details other than those about which distorted memories have been formed to count as reliably producing true beliefs.<sup>xviii</sup> Psychological findings about the misinformation effect establish that episodic memory systems are fallible but can also be viewed as undermining the claim that they are unreliable. Therefore, the argument of the current paper provides new support for the view that human memory systems could be reliable in spite of evidence of errors such as those

involved with the misinformation effect (Michaelian 2011, 2013), and certainly more reliable than they are often taken to be by psychologists who criticise eyewitness testimony (ibid.). Unlike existing arguments, however, it defends the surprising view that some eyewitness testimony is assigned less credibility than it should be given its reliability, even though people generally find the testimony to be persuasive and compelling.

## 10. Conclusion

Findings from cognitive science highlighting the susceptibility of human memory to the misinformation effect provide reason to worry that jurors, who tend to find eyewitness testimony compelling and persuasive, give too much credence to the testimony. This point has been widely recognised. However, this paper argues that the same psychological findings provide reason to worry that jurors who notice errors in a particular piece of eyewitness testimony will inappropriately lower the credence that they give to the testimony. They could respond to evidence of errors that occur due to the misinformation effect by concluding that the eyewitness who makes the errors lacks a supply of true beliefs that they are motivated to supply in their testimony. In such cases, jurors can inappropriately lower the credence that they give to some of the eyewitness testimony because it is consistent with eyewitnesses making errors in their testimony that they have a good supply of true beliefs that they are motivated to supply. In fact, evidence of errors in testimony can indicate the ordinary operation of cognitive mechanisms that facilitate eyewitnesses having a good supply of true beliefs that they can provide through their testimony.

In order to ensure that jurors are not susceptible to this error, they can be educated about the cognitive science literature that shows that people can make errors in their testimony while nonetheless having a good supply of true beliefs that they are motivated to provide. This literature can be used to guide jurors to make informed judgements about whether an error is due to a motivation to deceive or a lack of a supply of true beliefs, or due to the ordinary operation of human cognitive mechanisms that are likely to have facilitated the eyewitness gaining a good supply of true beliefs. There are currently debates about how to best achieve the goal of getting jurors to respond to evidence of this type, but the current discussion highlights that if a successful method is achieved, it will be possible to ensure that jurors do not lower the credence that they give to eyewitness testimony inappropriately.

## References

- Addis, D. R., Sacchetti, D. C., Ally, B. A., Budson, A. E., & Schacter, D. L. (2009b). "Episodic simulation of future events is impaired in mild Alzheimer's disease." *Neuropsychologia*, 47: 2660–2671.
- Alfano, M. (2013). *Character as moral fiction*. Cambridge University Press.
- Assefi, S.L. and Garry, M. (2002). "Absolute memory distortions: Alcohol placebos influence the misinformation effect." *Psychological Science*, 14: 77–80.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

- Belli R.F. and Loftus E.F. (1994). "Recovered memories of childhood abuse: a source monitoring perspective" in S.J. Lynn and J.W. Rhue (eds.) *Dissociation: Clinical and Theoretical Perspectives*, Guilford Press: New York; 415–433.
- Berman, G. L., & Cutler, B. L. (1996). "Effects of inconsistencies in eyewitness testimony on mock-juror decision making." *Journal of Applied Psychology*, *81*: 170-177.
- Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). "Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts." *Law and Human Behavior*, *19*: 79-88.
- Borckardt, J. J., Sprohge, E., & Nash, M. (2003). "Effects of the inclusion and refutation of peripheral details on eyewitness credibility." *Journal of Applied Social Psychology*, *33*: 2187-2197.
- Brainerd, C. J., & Reyna, V. F. (2002). "Fuzzy-trace theory and false memory." *Current Directions in Psychological Science*, *11*, 5: 164-169.
- Brigham, J.C. & Bothwell, R.K. (1983). "The ability of prospective jurors to estimate the accuracy of eyewitness identifications." *Law and Human Behavior*, *7*, 1: 19-30.
- Busby, J., & Suddendorf, T. (2005). Recalling yesterday and predicting tomorrow. *Cognitive Development*, *20*, 3: 362-372.
- D'Argembeau, A., & van der Linden, M. (2004). "Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance". *Consciousness & Cognition*, *13*: 844–858
- De Brigard, F. (2014). "Is memory for remembering? Recollection as a form of episodic hypothetical thinking." *Synthese* *191*, 2: 155-185.
- De Brigard, F., & Giovanello, K. S. (2012). "Influence of outcome valence in the subjective experience of episodic past, future and counterfactual thinking." *Consciousness and Cognition*, *21*(3), 1085–1096.
- Deese, J. (1959). "On the prediction of occurrence of particular verbal intrusions in immediate recall." *Journal of experimental psychology*, *58*, 1: 17-22.
- Bond, C.F. & DePaulo, B. M. (2006). "Accuracy of deception judgments." *Personality and Social Psychology Review*, *10*: 214–234.
- Dewhurst, S. A., Thorley, C., Hammond, E. R., & Ormerod, T. C. (2011). Convergent, but not divergent, thinking predicts susceptibility to associative memory illusions. *Personality and Individual Differences*, *51*(1), 73-76.
- Dodd, D. H., & Bradshaw, J. M. (1980). "Leading questions and memory: Pragmatic constraints." *Journal of Memory and Language*, *19*, 6, 695.

- Ekman, P., & O'Sullivan, M. (1991). "Who can catch a liar?" *American Psychologist*, 46: 913–920.
- Gabbert, F., Memon, A., Allan, K., & Wright, D. B. (2004). "Say it to my face: Examining the effects of socially encountered misinformation." *Legal and Criminological Psychology*, 9, 2: 215-227.
- German, T., & Nichols, S. (2003). "Children's counterfactual inferences about long and short causal chains." *Developmental Science*, 6: 514–523.
- Goldman, A. I. (1979). "What is justified belief?." In *Justification and knowledge* (pp. 1-23). Springer Netherlands.
- Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). "Patients with hippocampal amnesia cannot imagine new experiences." *Proceedings of the National Academy of Sciences of the United States of America*, 104: 1726–1731.
- Hatvany, N. & Strack, F. (1980). "The Impact of Discredited Key Witness", *Journal of Applied Social Psychology*, 10: 490-509.
- Howe, M.L. & Knott, L.M. (2015). "The fallibility of memory in judicial processes: Lessons from the past and their modern consequences." *Memory* 23, 5: 633-656.
- Howe, M. L., Garner, S. R., Dewhurst, S. A., & Ball, L. J. (2010). "Can false memories prime problem solutions?" *Cognition*, 117, 2: 176–181.
- Johnson, M. K. (1988). "Discriminating the origin of information" in T. F. Oltmanns & B. A. Maher (eds.), *Delusional beliefs* (pp. 34–65). New York: Wiley.
- Klein, S. (2013). The temporal orientation of memory: It's time for a change of direction. *Journal of Research in Applied Memory and Cognition*, 2, 222–234.
- Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). "Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future." *Social Cognition*, 20, 353–379.
- Laney, C. & Loftus, E. F. (2017). "Eyewitness testimony and memory biases." In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology*. Champaign, IL: DEF publishers. DOI:[nobaproject.com](https://doi.org/10.21963/noba.ch04.01)
- Lindsay D.S. (1994). "Memory source monitoring and eyewitness testimony" in D.F. Ross, J.D. Read and M.P. Toglia MP (eds). *Adult Eyewitness Testimony: Current Trends and Developments*, Cambridge University Press: New York, 27–55.
- Loftus, E. F. (2005) "Planting misinformation in the human mind: A 30-year investigation of the malleability of memory." *Learning and Memory*, 12: 361-366.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). "Semantic integration of verbal information into a visual memory." *Journal of experimental psychology: Human learning and memory*, 4, 1: 19-31.

- Loftus, E.F. & Palmer, J.C. (1974). "Reconstruction of automobile destruction: An example of the interaction between language and memory." *Journal of Verbal Learning and Verbal Behaviour*, 13, 5: 585-589.
- Meissner, C. A., & Kassin, S. M. (2004). "'You're guilty, so just confess!'" Cognitive and confirmational biases in the interrogation room.' In G. D. Lassiter (Ed.), *Interrogations, Confessions, and Entrapment* (pp. 85–106). New York: luer.
- Melo, B., Winocur, G., & Moscovitch, M. (1999). "False recall and false recognition: An examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures." *Cognitive Neuropsychology*, 16: 343–359.
- Michaelian, K. (2011). "Generative memory." *Philosophical psychology*, 24, 3: 323-342.
- Michaelian, K. (2013). The information effect: Constructive memory, testimony, and epistemic luck. *Synthese*, 190(12), 2429-2456.
- Michaelian, K. (2016a). *Mental time travel: Episodic memory and our knowledge of the personal past*. MIT Press.
- Michaelian, K. (2016b). Confabulating, misremembering, relearning: The simulation theory of memory and unsuccessful remembering. *Frontiers in psychology*, 7, 1857.
- Mitchell K.J. & Johnson M.K. (2000). "Source monitoring: attributing mental experiences" in E. Tulving & F.I.M. Craik (eds). *The Oxford Handbook of Memory* Oxford University Press: New York, 179–195.
- Mitchell, K. J., & Zaragoza, M. S. (1996). "Repeated exposure to suggestion and false memory: The role of contextual variability." *Journal of Memory and Language*, 35, 2: 246-260.
- Montmarquet, J. (1987). "Epistemic Virtue", *Mind*, 96: 482-97.
- Montmarquet, J. (1993). *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman & Littlefield.
- Neisser, U. (1967). *Cognitive Psychology*. New York, NY: Appleton.
- Okado, Y. and Stark, C.E.L. (2005). "Neural activity during encoding predicts false memories created by misinformation." *Learning and Memory* 12: 3–11.
- Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., et al. (2003). "Thinking of the future and the past: The roles of the frontal pole and the medial temporal lobes." *Neuroimage*, 19: 1369–1380.
- Puddifoot, K., & Bortolotti, L. (2018). Epistemic innocence and the production of false memory beliefs. *Philosophical Studies*, 1-26.

- Robins, Sarah K. (2016). "Misremembering." *Philosophical Psychology* 29, 3: 432-447.
- Roediger, H. L., & McDermott, K. B. (1995). "Creating false memories: Remembering words not presented in lists." *Journal of experimental psychology: Learning, Memory, and Cognition*, 21, 4: 803-814.
- Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). "Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives" in K.C. *Neuropsychologia*, 47: 2181–2187.
- Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: conceptual and perceptual false recognition. *Neuropsychology*, 11, 3: 331.
- Schacter, D. L., Verfaillie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language*, 35, 319–334.
- Schacter, D. L., & Addis, D. R. (2007). "The cognitive neuroscience of constructive memory: Remembering the past and imagining the future." *Philosophical Transactions of the Royal Society B*, 362: 773–786.
- Schacter, Daniel L., Addis, D.A. and Buckner, R.L. (2007) "Remembering the past to imagine the future: the prospective brain." *Nature reviews. Neuroscience* 8, 9: 657.
- Schacter, D. L., Guerin, S. A., and St. Jacques, P. L. (2011). "Memory distortion: an adaptive perspective." *Trends in Cognitive Sciences*, 15, 10: 467–474.  
<http://doi.org/10.1016/j.tics.2011.08.004>
- Schmechel, R. S., O'Toole, T. P., Easterly, C., & Loftus, E. F. (2006). "Testing Jurors' Understanding of Eyewitness Reliability Evidence," *Jurimetrics*, 46: 177-214.
- Seniuk, G. T. (2013). "Credibility Assessment, Common Law Trials and Fuzzy Logic." In *Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment* (pp. 19-30). Springer New York.
- Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 4: 527-538.
- Smith, V. L., & Ellsworth, P. C. (1987). "The social psychology of eyewitness accuracy: Misleading questions and communicator expertise." *Journal of Applied Psychology*, 72, 2: 294-300.
- Stein, E. (2003). "The admissibility of expert testimony about cognitive science research on eyewitness identification." *Law, Probability and Risk*, 2, 4: 295-303.
- Suddendorf, T. & Corballis, M. C. (1997) "Mental time travel and the evolution of the human mind." *Genetic Social and General Psychology Monographs* 123:133–167.  
<http://cogprints.org/725/>



Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel and is it unique to humans? *Behavioral and Brain Sciences*, 30: 299–313.

Sutton, J. (1998) *Philosophy and memory traces: Descartes to connectionism*. Cambridge University Press.

ten Brinke, L., & Porter, S. (2013). “Discovering deceit: Applying laboratory and field research in the search for truthful and deceptive behavior.” In *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 221-237). Springer New York.

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). “Discrepancy detection and vulnerability to misleading postevent information.” *Memory & Cognition*, 14(4), 329-338.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1–12.

Underwood, J., & Pezdek, K. (1998). “Memory suggestibility as an example of the sleeper effect.” *Psychonomic Bulletin & Review*, 5, 3: 449-453.

Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., & Van Overwalle, F. (2012). Counterfactual thinking: an fMRI study on changing the past for a better future. *Social cognitive and affective neuroscience*, 8(5), 556-564.

Vornik, L., Sharman, S., & Garry, M. (2003). “The power of the spoken word: Sociolinguistic cues influence the misinformation effect.” *Memory*, 11, 1: 101-109.

Williams, J. M., Ellis, N. C., Tyers, C., Healy, H., Rose, G., & MacLeod, A. K. (1996). “The specificity of autobiographical memory and imageability of the future.” *Memory and Cognition*, 24, 116–125.

Wilson, A. E. and Ross, M. (2003). “The identity function of autobiographical memory: Time is on our side.” *Memory* 11, 2: 137–49.

Wise, R.A., Fishman, C.A. and Safer, M.A. (2009). “How to Analyze the Accuracy of Eyewitness Testimony in a Criminal Case.” *Connecticut Law Review*, 42, 2: 435-513.

Zagzebski, L. (1996). *Virtues of the Mind* (Cambridge: Cambridge University Press)

Zaragoza, M. S., & Lane, S. M. (1998). “Processing resources and eyewitness suggestibility.” *Legal and Criminological Psychology*, 3, 2: 305-320.

Zaragoza, M. S., & Mitchell, K. J. (1996). “Repeated exposure to suggestion and the creation of false memories.” *Psychological Science*, 7, 5: 294-300.

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2007). “Misinformation effects and the suggestibility of eyewitness memory.” *Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom*, 35-63.

---

<sup>i</sup> I focus on eyewitness testimony for the sake of the current discussion, however, many of the arguments developed in the current article would also apply to testimony provided by complainants.

<sup>ii</sup> Other distorting factors have also been discussed in the psychology literature, including hindsight bias, eyewitness overconfidence, and tendency of eyewitnesses to make a relative judgement when identifying suspects in lineups (Wise et al 2009). To the extent that these effects are the result of the reconstructive nature of memory, the argument in this paper applies to them too.

<sup>iii</sup> Numerous articles have been written advocating jurors being informed by psychological findings on memory biases and responding to the view that eyewitnesses do not need to be informed by the findings because, for example, they are already aware of the fallibility of memory. Some examples include Stein 2003, Wise et al 2009 and Laney and Loftus 2017.

<sup>iv</sup> It is important to note that there is a significant epistemic flaw when incorporation occurs, even if the information that is incorporated is accurate: the believer and others have a false idea about the source of the memory belief. In some contexts, what people want from a testifier is evidence that they specifically witnessed some details of an event. For example, the testimony of one witness might be used to corroborate the testimony of another. If the corroborating testimony is the result of incorporation, especially the incorporation of information provided by the other witness, then the testimony will not provide the independent verification that it is taken to provide. In addition to this, sometimes testimony is taken to be especially valuable because it is the result of the personal experience of an eyewitness who is taken to be trustworthy. If the testimony is the result of the incorporation of information provided by another witness then the testimony will not perform the desired result. Each of these considerations suggest that although memory beliefs that involve incorporation will often accurately reflect details of remembered events, they will often nonetheless be misleading in ways that undermine them as sources of information in a criminal court. For current purposes, however, these worries shall be put aside because my positive argument does not depend on eyewitness testimony being reliable for the reasons that Michaelian (2013) outlines.

<sup>v</sup> In the current discussion it will not be possible to establish the frequency with which these cases occur. But even if I do not establish that the cases are common, it is important and notable that they occur, because they can lead to injustice, i.e. incorrect verdicts being made in criminal trials due to testimony being given less credence than it deserves.

<sup>vi</sup> Whether or not the eyewitness can be displaying intellectual virtue in the formation of the false memory will depend upon whether or not one is committed to the view that being virtuous requires success in attaining the truth (cf. Zagzebski (1996) and Montmarquet (1987; 1993)). All that is important for current purposes, though, is that the eyewitness is not displaying negative intellectual character traits that are likely to manifest in the formation of other beliefs, leading to other errors in testimony.

<sup>vii</sup> Recent discussions have questioned whether there are stable intellectual character traits that can count as intellectual virtues (Alfano 2013). I put this worry aside for the moment. All that is required for the current discussion is that it is possible that the eyewitness is behaving, in forming beliefs and recalling information about an event, in a way that could fittingly be described as intellectually virtuous (or not as intellectually vicious) if it were the expression of stable intellectual character traits.

<sup>viii</sup> To be clear, evidence of the misinformation effect is robust, suggesting that it occurs quite commonly. However, the claim is that source monitoring is still generally quite reliable, so we can trust that a person is often correct about the source of their beliefs. It is therefore incorrect to mistrust them about all of the information that they provide.

<sup>ix</sup> Stronger and weaker versions of this argument are possible. It could be argued that remembering is at least one of the primary function of the cognitive mechanisms that produce memories, but that the systems support other functions. Alternatively, it could be argued that remembering is not the primary function of the cognitive mechanisms that produce memory (see De Brigard 2014 for a version of the latter view).

<sup>x</sup> Because the episodic constructive simulation hypothesis shares with the mental time travel the claim that the function of the cognitive mechanisms underpinning memory is to facilitate prediction of the future it finds support in the abovementioned empirical findings supporting the mental time travel view that show that there are, e.g., similar cognitive underpinnings for judgements about the past and future and correlations between cognitive deficits in remembering the past and predicting the future.

---

<sup>xi</sup> De Brigard's view can be interpreted as a form of simulationism (Michaelian 2016b) or as an alternative to simulationism (Robins 2016). This matter can be put aside for current purposes.

<sup>xii</sup> Sarah Robins (2016) defends a *hybrid account of memory*, which integrates elements of constructivism with the traditional archival view of memory. Robins argues that it is only possible to explain the memory errors studied found in the DRM experiments if one accepts that memory systems retain traces of information about past events that are used in the process of constructing representations of the past. It is only possible to explain why participants often claim that they had studied words that were not on the list studied but were semantically related to the list words by supposing that the participants retain traces of information about the experience of reading the list. Therefore, remembering involves forming a representation of a past event through a combination of retention of information about the event and construction.

The distinguishing feature of the hybrid account is that it emphasises the importance of the retention of information. However, the account allows that construction is a central part of the process of memory formation. The view is therefore consistent with the cognitive mechanisms that produce memory beliefs performing various functions other than the retention of key information, such as the simulation of future events or the formation of plausible hypotheses about what might or might have happened.

<sup>xiii</sup> The information could be provided through instructions by a judge, by expert witnesses, through pre-trial training of jurors, or utilizing some other method found to be effective. For example, in the UK the Turnbull Guidelines laid down by the Court of Appeal in *Turnbull* [1977] QB 224 describe how when a case depends upon one or more identification of the accused, jurors should be instructed to consider ways that the eyewitness might be wrong. My argument provides reason for thinking the judge should include information about how the eyewitness might be right in spite of making errors.

<sup>xiv</sup> If representatives from either the defense or prosecution were required to deliver the information then one might doubt whether the information would be conveyed in an unbiased manner. However, from note 7 it can be seen that there are numerous ways that the relevant information could be delivered that would not involve either prosecution or defense representatives to deliver the information to jurors.

<sup>xv</sup> By advocating informing of jurors about the misinformation effect I do not deny the importance of adopting other strategies to prevent the misinformation effect, such as training police officers to avoid suggestive questioning. See, for example, "Code D: The Code of Practice for the Identification of Persons by Police Officers", point 3.3. where it is stipulated that a witness must not be shown images of a suspect before a video identification, identification parade or group identification. However, even if police officers were able to fully avoid suggestive practices, people would still be susceptible to the misinformation effect through social contagion effects, so it would still be of use for jurors to be informed about the effect.

<sup>xvi</sup> See note iii.

<sup>xvii</sup> Note: the juror would be engaging in source monitoring, but they would be monitoring another person's (i.e. the eyewitness's) memory rather than their own. They would therefore not be susceptible to the ordinary errors associated with monitoring of the source of one's own memory beliefs, such as those seemingly responsible for the misinformation effect.