Understanding the effects of lying by fabrication on memory for the truth

by

Harriet Rowthorn

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Psychology

University of Warwick, Department of Psychology

March 2020

Contents

List of Tables

List of Figures

Acknowledgements

First, I'd like to thank my supervisors, Kim Wade and Derrick Watson, aka 'the Dream Team'. Thank you, Kim, for the unwavering concern and support. I often lost belief in my ability and only recovered it thanks to Kim's constant faith in me. Thank you for inspiring me, leading by the highest example, providing incredible opportunities, and pushing for excellent scientific standards. Thank you to my second supervisor, Derrick, for the many hours of supervision and willingness to help me with any problem at any time. Every discussion left me wiser and reassured. I have the utmost respect and warmth towards both of you and I also want to thank you both for always making time for me, despite being so short of it yourselves. It was never unnoticed or unappreciated.

A special thank you to my fellow deception researcher, Danni, who could not have been a better friend and I feel eternally indebted to you. A further thanks to Divya and the remaining H1.24 office clan for the laughs, support and idiosyncrasies that make you all lovable. Meeting people like you was the best thing about doing a PhD.

Finally, I'd like to thank Warwick University and everybody there who works so hard to make it such a wonderful place to live and study. I am also so grateful to the ESRC for quite literally investing in me and giving me the opportunity to develop into a professional researcher. I'm excited to take the skills I have learnt to give something good back to the world.

Declaration

This thesis is submitted to the University of Warwick for the degree of Doctor of Philosophy. It is the result of the author's own work and has not been submitted for a degree or any other qualification at the University of Warwick or another institution. The work presented was executed solely by the author.

Abstract

Accurate memory matters nowhere more than in criminal cases, but research shows that people's memory reports can be contaminated by misinformation spread by others. An increasing body of evidence shows that people can also contaminate their own memories when they lie. However, a theory of how lying can affect memory has been slow to develop. The experiments within this thesis therefore aimed to further our understanding of how fabrication can affect later memory for the truth and test some of the hypotheses generated by two existing relevant theories: *von Hippel and Trivers' theory of self-deception* and the *memory and deception (MAD) framework*.

Experiments 1-3 test von Hippel and Trivers' proposal that repeating lies can promote forgetting of the truth via *retrieval-induced forgetting*. No evidence was found for this claim: Memory for information that participants repeatedly lied about did not differ from baseline. The thesis then turns to the relationship between the cognitive effort of fabrication and subsequent memory errors, as both theories are centered on the concept of cognitive load in deception. Some challenges are presented to the MAD framework's hypotheses concerning the relationship between the cognitive effort of fabrication and subsequent memory errors. The findings of Experiments 4-5 highlight the need to reconsider the nature of this relationship: Different methods of increasing liars' cognitive load moderated the effect of fabrication on memory. Finally, Experiment 6 tested the possibility that fabricating in a previous interview might decrease the accuracy of the Concealed Information Test (CIT)—a physiological deception detection test that relies on strong memory for the truth. Prior fabrication did not decrease the CIT's accuracy. Together, these findings help to refine models of lying and memory. Additionally, the findings raise concerns about contamination of revised testimonies after fabrication and highlight potential issues for increasing cognitive load for lie detection.

Chapter 1: Literature review

"If you want to keep a secret, you must also hide it from yourself"
– George Orwell (1954)

Deception is one of the greatest obstacles in the pursuit of criminal justice. Every statement that contributes to a criminal case is potentially a deceptive one: Offenders lie to escape incarceration, informants may be incentivized to provide false testimony (Informing justice, n.d.) and alibis can be false (Marion & Burke, 2013). Even those with a vested interest in the truth, such as victims and witnesses, may give false statements if they are intimidated (Lykken, 1998; "Time to Tackle", 2003). But despite its prevalence, deception is hard to detect. Most people perform little better than chance when trying to distinguish liars from truth tellers (DePaulo et al., 2003; The Global Deception Research Team, 2006) and the recurring phrase that there is "no Pinocchio's nose" makes for a pessimistic outlook on our ability to detect deception (Luke, 2019). Moreover, our ability to detect deception might be further hampered if a liar can come to believe their own lies, so that they not only deceive another, but potentially themselves. As Orwell knew, a secret is best kept when it is hidden from yourself.

One way that people may come to believe their own lies is through memory distortion. By feigning ignorance or fabricating false details, it is possible that the individual may forget truthful details or remember them incorrectly at a later date. Memory impairments induced by deception may be problematic in the criminal justice system for two reasons: First, poorer memory for the truth may indeed make it harder to detect when somebody is lying. Second, there may be instances where an individual wishes to retract a deceptive statement, such as when a witness enters protection or a suspect enters a plea bargain. In such cases, it is important that the individual can retrieve the truth to provide an accurate revised testimony. Yet there is relatively little research investigating the effects of lying on our ability to recall the truth, and even less research that contributes to our theoretical understanding of when and how memory is affected. This thesis aims to address this gap by testing some of the mechanisms that have been proposed in the existing literature.

1.1 Misinformation and memory

Lies can be considered a species of misinformation in the sense that they convey false or inaccurate information. We already know from decades of misinformation research that exposing people to inaccurate information about an experienced event can lead them to later recall that event incorrectly, either partially or entirely (Loftus, 2005; Wade, Rowthorn, & Sukumar, 2017). Even without exposure to false information, memory for an event can be affected simply by the way in which a story is recounted. As far back as 1932, Bartlett highlighted that memory is not merely reconstructive, but a social act: We typically report only relevant aspects of an event and omit other details depending on the audience and the goals of the exchange. This selective reporting also influences what details are later remembered about the original event (Tversky & Marsh, 2000).

When people lie, they typically either mislead the receiver by fabricating false information or omit information that the receiver wishes to know. A deceptive individual therefore often creates their own post-event misinformation or selectively reports information about that event. How might this affect that individual's memory? One study showed that both introducing false information and omitting true information can have powerful effects on memory (Wright, Loftus, & Hall, 2001). Participants watched a slide sequence depicting a restaurant scenario and then retold the sequence of events they had seen by using story cards provided by the experimenter. Some participants were given a set of cards containing an extra event that did not happen in the original scene, whereas other participants were given an incomplete set of cards that omitted an event from the original scene. Some participants therefore had the opportunity to include in their retelling an event that did not happen (as they would if they were fabricating) whereas others were forced to omit from their retelling a critical event that did happen (and therefore selectively reported the event).

When participants' memories were tested one week later, 20% of participants who included the extra false event in their retelling recalled this incorrect detail compared to 6% of the control group who did not retell the scenario using story cards. Additionally, only 44% of participants who omitted the critical event from their retellings recalled this detail, compared to 67% of the control group (Wright et al., 2001). Embellishing stories with false information therefore increased the likelihood that participants mistook false information for the truth, whereas omitting critical details from their stories increased the likelihood that participants would forget this detail compared to if they had not rehearsed the story at all.

Although classic misinformation studies like Wright et al. (2001) show that misinformation and selective reporting can distort memory, the findings do not necessarily generalize to deception. The participants in these experiments were not deliberately embellishing or omitting details from their story and may have been unaware that they were doing so. However, a related body of research suggests that people remain vulnerable to misinformation effects even when they knowingly generate that misinformation themselves. Instead of exposing participants to post-event misinformation, these studies force participants to generate their own misinformation by asking them ‘impossible’ questions about an event they witnessed; that is, questions that have no corresponding answer because they address details that were never shown. For instance, after witnessing a child fall from a chair, participants might be asked where he was bleeding from, despite the fact that he was shown unharmed (Ackil & Zaragoza, 2011). Participants are therefore forced to fabricate a response to complete the task.

These “forced fabrication” studies are designed to mimic coercive interview settings where an eyewitness is made to guess or speculate about a detail they cannot remember. The research to date consistently shows that participants often falsely remember the details that they were forced to fabricate in both recall and recognition tests (Ackil & Zaragoza, 1998; Pezdek, Sperry, & Owens, 2007; Zaragoza, Payment, Ackil, Drivdahl, & Beck, 2001; Zaragoza, Rich, Rindal, & DeFranco, 2017). People are therefore susceptible to incorporating misinformation into memory, even when they have generated that misinformation themselves. Perhaps most puzzling about this effect is that participants develop false memories of the details they were forced to fabricate despite often strongly resisting the task of answering the questions. Indeed, people can come to falsely remember not only small details, but entire fabricated events: In one study, participants developed false event memories in 50% of cases over an 8-week period (Chrobak & Zaragoza, 2008).

Although research on forced fabrication is informative, forced fabrication differs from deliberate fabrication in a number of important ways. Participants in forced fabrication studies are not informed that some of the questions they will be asked have no corresponding answer. They might therefore reasonably assume that they are being asked about events that did in fact occur and the belief that those events occurred might make participants more susceptible to memory distortion. This is consistent with the finding that factors that increase participants' uncertainty in their fabricated responses (such as higher resistance to providing an answer or receiving a warning that they will be asked about false events) decrease false memory rates (Ackil & Zaragoza, 2011). In other words, when participants are more confident that the event in question never occurred, they are later less likely to falsely recall their fabricated details. In deliberate fabrication, participants are indeed likely to be confident that the event did not occur because the purpose of their task is to communicate false information and it is important for a liar to be aware of which information is true and which false. The intent to deceive also requires deliberate fabricators to *appear* certain in their responses so that they are more likely to be believed. Deliberate fabricators might therefore be less prone to memory distortion than forced fabricators.

Another important difference between forced and deliberate fabrication is that the details participants are forced to fabricate are often consistent with the facts of the event, whereas deliberate fabrications may instead contradict the truth. To take the above example, the suggestion that the boy was bleeding is consistent with the witnessed event of him falling from a chair. Conversely, a deliberate fabricator might instead replace a truthful detail with a false one. For instance, a criminal suspect might say that they arrived home at 4pm on the night of the crime, rather than at 11pm. People may be more resistant to developing false memories from such contradictory fabrications, since other supporting memories of the event may be used as evidence in the memory verification process to reject the false information as fabricated (Wade & Garry, 2005). Consistent with this idea, people are more prone to developing false memories of information they have been forced to fabricate when that information explains how or why the witnessed events occurred (Chrobak & Zaragoza, 2013). This suggests that the consistency of the fabricated information with the event facts could be an important factor for false memory development.

It is therefore plausible that deliberate deception might have different effects on memory than externally sourced misinformation or misinformation that one has been forced to fabricate. Lying is a unique category of misinformation, defined as "a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue" (Vrij, 2008, p. 15). A liar therefore knows that what they are saying is false and because they must have the intention to create a false belief, lying is likely to be a salient act for the individual. Indeed, existing evidence on how lying affects memory is mixed, with some studies showing that lying can impair memory for the truth (e.g., Pickel, 2004; Polage, 2012), while others show that lying has no unique effects on memory (e.g., Vieira & Lane, 2013), or can even enhance it (e.g., Polage, 2004). However, there is little research that synthesizes our existing knowledge to make sense of when memory is impaired, unaffected or enhanced by lying. Such inconsistent findings suggest that lying may indeed be a distinct category of misinformation and therefore its effects on memory warrant more detailed investigation.

One factor that distinguishes lying from other kinds of misinformation is that liars play an active role in the spread of false information by intending to induce a false belief in another and, in some cases, generating that false information themselves. This means that they are more likely to experience greater cognitive load than an individual who passively receives misinformation from an external source (for example, the media, or an experimenter). The relationship between cognitive load and lying has been offered as a potentially important factor for understanding when, how and why lying affects memory for the truth, which is discussed in the forthcoming section.

1.2 Lying and cognitive load

The common denominator of almost all deception research is the notion that lying generally requires more cognitive resources than truth telling, and that this places liars under greater cognitive load (Verschuere, Köbis, Bereby-Meyer, Rand, & Shalvi, 2018; Vrij, Fisher, Mann, & Leal, 2008; Vrij, Fisher, & Blank, 2017; Walczyk, Harris, Duck, & Mulay, 2014). “Cognitive resources” refers to the “limited pools of attention and working memory available to respondents” (Walczyk et al., 2014, p.23) and “cognitive load” is the extent to which these resources are depleted. Put simply, lying typically requires more processes than truth telling.

What makes lying typically more effortful than truth telling? Truth tellers can simply retrieve and relay the truth, whereas liars must organize, manipulate and elaborate on existing memories or scripts to generate false counterfactual information that is consistent with the known facts (Sporer, 2016). Not only is constructing the lie itself cognitively demanding, but also delivering the lie. The cognitive demand of lie delivery stems from two main factors: Monitoring and suppression (Vrij, Fisher, et al., 2008). Unlike truth tellers, liars do not assume that they will be believed and they therefore must monitor their behavior to ensure that they do not leak cues to dishonesty. Liars must also monitor the behavior of the enquirer to assess if they are believed (Schweitzer, Brodt, & Croson, 2002). Such monitoring requires cognitive resources. Additionally, because the truth is automatically activated (when accessible), liars must dedicate resources to the deliberate activation of a lie (Duran, Dale, & McNamara, 2010) and the suppression of the truthful response (Spence et al., 2001) at the time of lie delivery.

One reason that the concept of cognitive load receives so much attention from deception researchers is that increased cognitive load often manifests in liars’ behavior and can be used to facilitate deception detection (Vrij, Fisher, et al., 2008). Indeed, people are more accurate at detecting deception when they are asked to look for signs that somebody is under increased cognitive load, as opposed to when they are asked to simply state whether they think that the individual is lying (Vrij, Granhag, Mann, & Leal, 2011; Walczyk et al., 2012).

Strategies that reduce the cognitive cost of lying may therefore help an individual to elude detection. Memory distortion may be one such strategy: If an individual comes to believe that their lies are the truth, they are no longer cognitively taxed by suppressing suspicious behaviors or the memory for the truth. In fact, a prominent theory of self-deception proposes that self-deception evolved for precisely this purpose, positing that people can come to believe their own lies as a mechanism for reducing the cognitive demand of lying and more effectively deceiving others (von Hippel & Trivers, 2011).

1.3 von Hippel and Trivers' theory of self-deception

According to von Hippel and Trivers' theory of self-deception, maintaining an accurate representation of the world would not have worked to our evolutionary advantage. Instead, the authors propose a co-evolutionary struggle between the deceiver and deceived, such that deception evolved to help individuals accrue resources, but this in turn selected for better detection strategies to prevent us from being duped so often. The key claim of the theory is that self-deception is the next step in this co-evolutionary battle: A deceptive individual is much harder to detect if they themselves come to believe the lie, since they should leak fewer cues to deception than somebody who knowingly communicates a falsehood. There are numerous types of self-deception proposed by von Hippel and Trivers' theory, each with different underlying mechanisms. Of relevance to the current thesis is their proposal that self-deception can be achieved via memory distortion.

One question is how such self-induced memory distortion could occur. People cannot simply choose what they do and do not believe, so if self-deception was functional in our evolutionary history, there must be specific mechanisms that enable people to maintain inaccurate views of the world, despite there being evidence to the contrary. Von Hippel and Trivers (2011) suggest two specific mechanisms that might promote memory distortion following deception: *Retrieval-induced forgetting* (RIF; Anderson, Bjork, & Bjork, 1994) and *source monitoring errors* (Johnson, Hashtroudi, & Lindsay, 1993). The experiments presented throughout this thesis test these mechanisms to investigate whether fabricating false information can indeed impair memory for the truth via RIF and source monitoring errors. These mechanisms, and the existing evidence that suggests they may be responsible for deception-induced changes in memory, are outlined in the next section.

1.4 Retrieval-induced forgetting (RIF)

Von Hippel and Trivers (2011) propose that by simply repeating a lie and omitting the truth, people may come to forget the truth via the process of RIF. Repeating lies may therefore promote *omission errors*; that is, fewer correct event details are recalled after lying about that event, which can also be described as forgetting the truth.

Though RIF has received little attention in deception research, it has been extensively studied in the broader memory literature (Murayama, Miyatsu, Buchli, & Storm, 2014; Storm et al., 2015). RIF simply refers to the finding that strengthening memory in one respect can impair it in another; memory is improved for practiced information, but impaired for related unpracticed information (Anderson et al., 1994). For instance, when people practice retrieving the item "orange" in response to "Fruit-Or__", they are later less likely to retrieve the unpracticed item "lemon" in response to "Fruit-Le__" than if they had not practiced retrieving the item "orange" (Anderson et al., 1994). The idea is that practicing lies might impair memory for the corresponding truthful information in a similar way.

RIF has been studied in a forensic context, specifically looking at the effects that selective retrieval practice might have on eyewitness memory. Given that eyewitnesses are likely to be questioned on only a subset of the details from a crime, they can experience RIF for other aspects of the crime that are omitted from their statements. For instance, an investigator might focus on certain important objects like a stolen purse, but neglect to ask about other objects that later become relevant to the investigation, like the culprit's backpack, and memory for these objects can suffer as a result (Storm et al., 2015).

Studies investigating RIF in an eyewitness context typically replace the category-item verbal stimuli like "Fruit-Orange" with mock crime scenarios in the form of descriptions, images or videos. After witnessing the mock crime, participants are questioned about a subset of the details they saw. Subsequent recall and recognition tests indicate that this selective questioning does indeed elicit RIF. In one study, for instance, participants imagined that they were police officers examining a crime scene and were shown images of items allegedly stolen from two houses. They were then questioned on half of the items witnessed from one of the houses. When participants were later tested on all of the items from both houses, their memory for items that they were not initially questioned on was significantly poorer when those items came from the same house as the items they were questioned on, compared to when those items came from the other house (MacLeod, 2002). Questioning a witness about a subset of what they experienced therefore led to RIF; that is, it impaired memory for related information that was not initially probed.

Existing research studying the effects of RIF in a forensic context has focused almost exclusively on eyewitnesses, showing that eyewitness memory for objects, offender characteristics and actions are all vulnerable to RIF (Camp & Wesstein, 2012; MacLeod, 2002; Migueles & García-Bajos, 2007; Shaw, Bjork, & Handal, 1995). From this research, we know that investigative interviewing can inadvertently promote forgetting of information that might prove relevant later in an investigation. This may be true regardless of whether the interviewee is responding honestly or deceptively. Just as a witness is unlikely to be exhaustively questioned about a crime they witnessed, a lying suspect is also unlikely to be exhaustively questioned about a crime they have committed. RIF may therefore be a problem in other investigative contexts to eyewitness memory. But what causes the memory impairments consistently shown in these studies?

The cause of RIF has been debated for more than 20 years (Buchli, Storm, & Bjork, 2016; Murayama et al., 2014; Storm et al., 2015), but the most recent meta-analysis concludes that the evidence favors an inhibition-based account overall (Murayama et al., 2014). According to the inhibitory account, attempting to retrieve a target item (e.g., “orange”) activates other non-target associated items (e.g., “lemon”) that share a retrieval cue (i.e., “fruit”). These non-targets compete with the target for retrieval, thereby creating interference, which is resolved by inhibiting non-targets. Repeated inhibition from multiple retrieval attempts reduces the accessibility of the inhibited items, which impairs memory for these items (Anderson et al., 1994; Murayama et al., 2014). In other words, practicing retrieval of an item promotes inhibition of related items in memory, which impairs later memory for those related items.

Although the inhibitory account has considerable evidence behind it, there is a rival explanation that has been proposed to underlie RIF at least in some circumstances—the competition-based account. According to the competition account, inhibition is unnecessary to explain the forgetting effects following selective retrieval practice. Instead, its proponents claim that practicing retrieval of a subset of items strengthens their representation in memory. These stronger memories then interfere with the retrieval of weaker, unpracticed items in memory and block their retrieval (Raaijmakers & Jakab, 2013). For example, practicing retrieving "orange" in response to "Fruit-Or__" strengthens the memory of the association between "orange" and "fruit" so that it blocks retrieval of weaker memories associated with the cue "fruit", like "lemon". Thus, "lemon" is less likely to be recalled simply because stronger memories interfere with its retrieval.

The critical difference between the inhibitory and competition-based accounts of RIF concerns their hypotheses about the strength of the memory for unpracticed items. The inhibitory account says that the strength of the memory for unpracticed items is directly weakened by an inhibitory process engaged during retrieval practice. The competition-based account instead says that the strength of the memory for unpracticed items is not weaker, but their retrieval is just temporarily blocked by stronger items in memory. Thus, RIF represents genuine forgetting only for the inhibitory account. For the competition-based account, RIF is not really forgetting at all, but merely a product of the way that memory is being tested (Chan, Erdman, & Davis, 2015).

Importantly, the memory impairments shown in RIF studies are not simply due to a lack of rehearsal and therefore do not merely reflect the passive decay of memories over time. RIF is a unique category of forgetting caused by rehearsing other information. We know this because people's memory for unpracticed items that are unrelated to those practiced is consistently better than their memory for related unpracticed items (Murayama et al., 2014). To continue with the example above, people's memory for unrelated unpracticed items such as “whisky” in response to “Drink-Wh__” is better than their memory for the related unpracticed item “lemon”. In other words, neither “lemon” nor “whisky” were practiced, yet memory for “lemon” suffers as a direct result of practicing “orange”, whereas memory for “whisky” does not. Memory for “lemon” is therefore below baseline: It is poorer than it would have been had the individual not practiced retrieving any items at all.

It is the “below baseline” forgetting that makes RIF distinct from mere decay over time. Experiments must therefore be designed using procedures that can differentiate between RIF and decay to establish that RIF is the mechanism underlying any forgetting effects found. While several studies investigating the effect of lying on memory have suggested RIF as a mechanism, this is typically offered as a post hoc potential explanation of forgetting effects (e.g., Christianson & Bylin, 1999; Gronau, Elber, Satran, Breska, & Ben-Shakhar, 2015; Paige, Fields, & Gutchess, 2018; Polage, 2018). Such experiments therefore do not include appropriate controls to determine that lying can indeed promote RIF of the truth. Thus, there is currently no direct research to support von Hippel and Trivers (2011) proposal that RIF could be one way that self-deception is achieved.

While there is little research that explicitly investigates RIF following deception, there is a body of research studying omission errors—that is, forgetting—caused by false denials (denying aspects of or the entirety of an event that did in fact occur) and feigned amnesia (pretending to have forgotten about aspects of or the entirety of an event that occurred). We therefore do have some knowledge of how lying could promote forgetting. An overview of this research is given below.

1.4.1 Deception-induced omission errors

1.4.1.1 False denials

Studies focused on false denials—that is, denying aspects of or the entirety of an event that in fact occurred—show that the most pronounced effect of denial on memory appears not to be for the details denied, but instead for the interview in which those details are discussed. In one experiment, participants were shown a series of images and were then asked a series of questions about details from a subset of those images. Participants responded in one of three ways: The false denial group were instructed to (wrongly) deny that the event or object in each question was shown. A second 'external denial' group were instead instructed to respond honestly, but received feedback from the experimenter that the event or object stated in their answer never occurred. The final group of participants instead responded honestly by truthfully reporting the details they remembered from the images. The next day, participants completed a recognition test and a source test to assess their memory for the details from all of the images they studied, as well as their memory for whether they spoke about each detail during interview. For all groups, there was no impairment to recognition memory; that is, memory for the image details addressed in the interview was not affected. However, source memory was poorer in the false denial group: False deniers were less likely to remember having discussed those details compared to the external denial and honest groups (Otgaar, Howe, Smeets, & Wang, 2016). Thus, falsely denying impaired memory for the interview, but not the details originally witnessed– a finding called denial-induced forgetting (DIF).

DIF is a non-trivial finding: Poorer memory for what was discussed during interview may have its own consequences for lie detection. An individual may be less consistent over repeated interviews if they struggle to recall how they have previously responded. Since inconsistency is often interpreted as a cue to deception or low reliability (Smeets, Candel, & Merckelbach, 2004; Vredeveldt, van Koppen, & Granhag, 2014), DIF could actually facilitate lie detection by promoting inconsistencies across interviews.

DIF has been consistently replicated across a range of stimulus types and in both recognition and recall (Otgaar, Howe, Memon, & Wang, 2014; Otgaar, Romeo, Ramakers, & Howe, 2017; Romeo, Otgaar, Smeets, Landstrom, & Boerboom, 2018). It should be noted, however, that DIF studies compare false deniers with truth tellers and do not include a control group who were not interviewed—given that DIF concerns memory for the interview, it is not possible to have a comparison group who are not interviewed. Nonetheless, this means that it cannot be determined if DIF represents a true forgetting effect arising from false denial or enhanced memory for the interview in truth tellers. Additionally, the lack of an appropriate control can obscure the mechanism underlying the forgetting effect, as it means that there is no measure of the individual's baseline forgetting rate. It is therefore difficult to determine if DIF is caused by a specific process that is recruited when denying, such as an inhibitory mechanism, or simply due to a lack of rehearsal.

While DIF is a robust finding, the effect of false denials on item memory (that is, memory for the details denied) is less consistent. The DIF studies cited above consistently show that item memory is unaffected. Furthermore, one experiment found that memory for previously studied details may even be improved after denying having studied them (Vieira & Lane, 2013). Nonetheless, there is some evidence suggesting that denial can impair memory for the details denied. Denying knowledge of crime-relevant details in a reaction-time-based lie detection test has been shown to increase response times to those details when the individual subsequently switches to truth telling. Participants were also less accurate when responding honestly after having lied (Visu-Petra, Jurje, Opre, Miclea, & Visu-Petra, 2014). Slower responses suggest that denying knowledge of the crime-relevant details decreased their accessibility in memory, however this finding could merely reflect the switch cost of alternating between the task of lying and truth telling (Mayr & Kliegl, 2000; Visu-Petra et al., 2014).

It is possible that false denials do in fact reduce the accessibility of denied items in memory, but that the effect is too subtle to detect with forced-choice recognition tests such as those used in DIF studies and Vieira and Lane (2013). This idea is supported by one study showing that participants who consistently denied the occurrence of a true childhood event reduced their belief that they experienced that event (Polage, 2018). Participants did not, however, move from belief to unbelief that the event occurred as a result of denying it. Thus, it may be that false denial slightly reduces the accessibility of denied items in memory, but not enough to promote forgetting and therefore omission errors do not significantly increase.

In sum, the evidence to date suggests that false denials impair source memory, such that people are worse at remembering when they denied details in a previous interview than they are at remembering when they told the truth. The effects of false denials on item memory is inconclusive, but the evidence suggests that if false denials do impair memory for the details denied, the impairment is likely to be too subtle to significantly increase omission errors beyond an individual's baseline rate of forgetting. Thus, the evidence currently favors the hypothesis that there is nothing special about false denials that promotes forgetting over and above the decay that would occur with the passage of time.

1.4.1.2 Feigned amnesia

The effect of feigning amnesia on memory has received more attention from researchers than the effect of false denials, primarily because feigning amnesia is fairly common: It has been estimated that 23% of violent criminals claim partial or total amnesia for the crime committed (Cima, Nijman, Merckelbach, Kremer, & Hollnack, 2004) and it is therefore important to know how this might affect memory. Experiments studying the effects of feigned amnesia on memory typically follow a similar procedure. Participants first read about or perform a mock crime before undergoing interview, in which some participants feign memory loss for the crime-relevant details. They then honestly complete a free and cued recall test to assess their memory for the crime-relevant details. This same test is repeated after a delay of approximately 1 week to determine how feigning amnesia has affected their memory for the crime.

In one of the first studies using this procedure, participants' memories actually improved across the two sessions after feigning amnesia, however their memory performance was still below that of participants who responded honestly. For honest responders, 45% of the details recalled were correct, whereas 37% of the details recalled were correct for participants who feigned amnesia (Christianson & Bylin, 1999). Subsequent research shows a similar pattern of findings (Van Oorsouw & Merckelbach, 2006). The authors of both studies conclude that feigning amnesia does indeed impair subsequent recall and propose that selectively reporting truthful details of the crime, while pretending to have forgotten others, might have promoted genuine forgetting of those details via RIF. It is worth noting that if RIF does underlie the memory undermining effects of feigned amnesia, it would not be feigning amnesia *per se* that drives the memory impairment, but the selective truthful retrieval of other crime-relevant information.

A recent study set out to directly test the hypothesis that RIF is the mechanism underlying the forgetting effects found in feigned amnesia studies. The authors adapted the classic retrieval practice paradigm from Anderson et al. (1994) such that participants practiced retrieving a subset of crime-relevant details and omitted others to determine if selective retrieval practice impairs memory for the omitted details (Mangiulli, Van Oorsouw, Curci, & Jelicic, 2019).

Participants watched a mock crime video and were asked to imagine that they were the perpetrator, who engages in a violent fight. They were then allocated to either a retrieval practice condition, control, or truthful recall condition. Participants in the retrieval practice condition studied a subset of the details from the crime and then wrote down these details to mimic a situation where an interviewee selectively reports some crime-relevant details, but omits others by claiming memory loss. This process is analogous to the retrieval practice phase of RIF experiments, where participants practice retrieving some items they have previously learnt, while omitting others. Participants in the control condition were instead asked to simulate memory loss for crime-relevant details, but did not receive any further instructions or prompting for which details to omit and report. On completion of their tasks, participants in both groups were then asked to honestly freely recall as many details from the crime as possible. Participants in the truthful condition progressed straight to this honest free recall task after experiencing the mock crime. Finally, all participants returned the next day to complete the free recall test for a second time (Mangiulli et al., 2019).

Regardless of whether memory was tested immediately or after a 24-hour delay, participants in the retrieval practice condition recalled fewer omitted details than did participants in the control condition. The authors interpret this finding as evidence that selectively retrieving a subset of crime-relevant details promoted RIF of the information participants omitted in their original reports. However, this interpretation is contestable. It is unclear that the procedure is an appropriately modified version of the retrieval practice paradigm to yield such a conclusion. A critical component of the retrieval practice paradigm is that memory for items that are selectively practiced and omitted is compared to a baseline measure of memory for items that are unrelated to those participants have practiced retrieving. This is the requisite comparison to conclude that RIF underlies any forgetting effects. However, the comparison group in Mangiulli et al. (2019) was the control group who feigned amnesia without specific instruction. These participants still therefore selectively retrieved and omitted crime-relevant information and are not therefore an appropriate comparison with the retrieval practice group. Additionally, participants in the retrieval practice group were not explicitly instructed to feign memory loss at all, whereas participants in the control group were. It might therefore be argued that participants in the retrieval practice group did not feign amnesia at all. These factors lead to interpretive issues: It is unknown whether the memory impairment found in the retrieval practice group is below baseline memory performance– which is a requirement of RIF– and it is also unclear whether the impairment reflects any processes that are related to feigned amnesia at all.

Since the aforementioned studies do not include a no-rehearsal control group who neither feigned amnesia nor responded honestly, it remains unclear whether feigning amnesia can impair memory. The lack of a no-rehearsal group means that these studies cannot provide evidence that feigning amnesia promotes forgetting over and above what we would find from the mere passage of time. In fact, studies that include a no-rehearsal control group typically find that the memory performance of participants who feigned amnesia does not differ from the no-rehearsal group, suggesting that feigning amnesia seems to reduce memory performance only because it prevents the truthful rehearsal of crime-relevant details (Mangiulli, Van Oorsouw, Curci, Merckelbach, & Jelicic, 2018; Sun, Punjabi, Greenberg, & Seamon, 2009; Van Oorsouw & Merckelbach, 2004).

In sum, the research to date provides little evidence that lying by false denials or feigning amnesia facilitates forgetting over and above what we would see from the mere passage of time. These types of lie do not appear to engage specific processes that increase omission errors beyond the baseline rate of forgetting. Thus, in the context of denial and feigned amnesia, there is currently no evidence to support von Hippel and Triver's (2011) hypothesis that RIF can underlie self-deception. However, research into the cognition of deception has revealed that there are specific processes recruited when people lie that may promote RIF of the truth following repeated fabrication. This is explored further in Chapter 2, in which the experiments reported investigate whether repeating fabrications can promote RIF of the truth via an inhibitory mechanism.

While the evidence for deception-induced omission errors is weak and inconsistent at best, the evidence for deception-induced commission errors is far stronger. This is discussed in the forthcoming section in relation to the second mechanism that von Hippel and Trivers (2011) propose to underlie self-deception: Source monitoring errors.

1.5 Source monitoring

According to von Hippel and Trivers (2011), a second way that people may come to believe their lies is by "self-inducing false memories" (von Hippel & Trivers, 2011, p. 6). While somebody may initially knowingly report a falsehood, over time they may confuse the source of the false information they have generated, such that they later believe it to be the truth. Fabricating false information could therefore increase commission errors; that is, it may lead people to report more incorrect details.

The source monitoring framework (SMF; Johnson et al., 1993) is the most widely used framework for explaining how people can develop false memories. According to the SMF, memories are not encoded with labels that tell us where they originated when we retrieve them at a later date. Instead, source monitoring is a decision-making process where we infer the source of our memories based on certain subjective qualities they have (Johnson, Foley, Suengas, & Raye, 1988). Memories from different sources are systematically different and we use these differences to decide whether our memory is based on a genuine experience or an internally generated one (for instance, a fabrication).

Genuine memories tend to be more vivid than internally generated memories. A genuine memory is typically richer in perceptual and contextual details, as well as meaning and emotional relevance (Johnson et al., 1993). For example, recalling a childhood trip to the seaside is likely to cue the full spectrum of your experience at the time: You may see in your mind's eye the waves crashing against the shore and recall the smell of the sea, the sound of gulls overhead and irritation at trying to rid yourself of the gritty sand between your toes. If you never had this experience, there will be fewer of these associated memories and therefore fewer indicators that this was something that you genuinely experienced.

If you had instead merely imagined a childhood seaside trip, not only will your memory have fewer contextual and perceptual details, but you are also likely to have more cognitive operations associated with your memory. Cognitive operations are the mental processes that are activated when encoding information (Foley & Foley, 2007). These include the retrieval processes that allow you to piece together memories from other relevant experiences, as well as your knowledge and expectations about what the seaside is like so that you can imagine that such a trip happened. Because these processes are relatively effortful, the presence of cognitive operations is an indicator that an event was not truly experienced, but internally generated.

The presence or absence of the aforementioned qualities therefore help us to judge whether a memory is authentic: Greater perceptual details and contextual embeddedness, together with few associated cognitive operations mean that we are more likely to attribute a memory to a genuine experience. Conversely, a less perceptually and contextually vivid memory, together with more cognitive operations mean that we are likely to attribute a memory to an internally generated experience, like a fabrication (Lindsay, 2014).

Most of our source monitoring occurs unconsciously and automatically, and is largely accurate (Lindsay, 2008). But errors do happen: If an authentic memory lacks vivid perceptual details and contextual embeddedness, it might be incorrectly attributed to an internally generated experience. Conversely, if an internally generated event is very rich in these details and is retrieved with little effort, such as a very vivid dream, it might be mistaken for an authentic memory. Such mistakes are called *source monitoring errors* (Johnson et al., 1993), but for consistency throughout this thesis I refer to them as *commission errors*.

Returning to deception, it is possible that liars can mistake their lies for the truth through failures of source monitoring, as von Hippel and Trivers (2011) suggest. A fabrication rich in perceptual and contextual detail may be mistaken for a genuine experience, particularly if there are few cognitive operations associated with the lie. Indeed, there is a growing body of research showing that fabricating and reporting false information can increase commission errors via failures in source monitoring. This literature is reviewed below.

1.5.1 Deception-induced commission errors

One of the first studies to investigate the effect of fabrication on memory looked at deliberate fabrication in an eyewitness context (Pickel, 2004). Participants watched a mock robbery video before undergoing interview or allocation to a control group with no interview. Participants undergoing interview either fabricated incorrect information about what they saw, as an intimidated witness might, or truthfully responded to questions. One week later, participants' memories for the original details of the crime were tested. Participants who fabricated information about the perpetrator recalled just as many correct details as control participants who were not interviewed, but they recalled significantly more incorrect details. Fabrication therefore increased commission errors, but not omission errors. Moreover, of the incorrectly recalled details, 27% were lies that participants had fabricated during interview. Thus, in over a quarter of cases, they confused their lies for the truth.

The more believable participants' lies were (as assessed by two judges), the more likely their fabrications were confused with the truth. Believable lies may be easier to visualize and therefore more perceptually vivid, increasing the likelihood that they will be judged as authentic memories. They are also more likely to be more consistent with witnesses' wider knowledge of the event and the greater similarity between the lie and truth may have made source monitoring harder (Pickel, 2004, Exp 1). Thus, lie plausibility appears to be one factor that moderates the extent to which people come to mistake their lies for the truth by virtue of making source monitoring more difficult.

In a follow up experiment, participants either generated their own fabrications, as in the previous experiment, or were given experimenter-generated fabrications to mimic a situation where an individual is provided with a false alibi or told what to say by another (Pickel, 2004, Exp 2). Regardless of whether the fabrication was self- or experimenter-generated, liars made more commission errors and more omission errors than truth tellers. Note, however, that the experiment lacked a control group who did not undergo interview, so we cannot know if the increased omission errors reflect a memory impairment from participants who lied or a memory enhancement from participants who reported the truth. Nonetheless, the rate of commission errors was affected by the source of the fabrication: Participants who reported experimenter-generated fabrications incorrectly recalled 56% of their fabrications from the interview as the truth, compared to 37% for subjects who generated their own fabrications (Pickel, 2004, Exp 2).

The finding that experimenter-generated fabrications are more likely to be incorporated into memory than self-generated fabrications has subsequently been replicated in the forced fabrication literature. In one study, participants witnessed a mock crime and were then forced to answer ‘impossible questions’ that addressed aspects of the crime that never in fact occurred. Some participants generated their own false answer, whereas others received a suggested false answer within the question that they could simply copy. Participants were more likely to incorporate the fabrications into memory when they were suggested in the question than when they generated them themselves (Pezdek, Lam, & Sperry, 2009).

Similarly, a more recent series of studies showed that participants were more likely to recall their fabricated answers to general knowledge questions than their truthful answers, but only when they had fabricated the answers themselves. When participants instead chose between the truthful answer and a predetermined fabrication from the experimenter, they no longer showed a memory advantage (Besken, 2018). Taken together, these findings suggest that a victim or witness who is forced to state incorrect details may be at greater risk of memory impairment than somebody who volunteers a lie.

The findings of these experiments are consistent with the SMF, which predicts that commission errors are more likely when a memory is rich in perceptual vividness and contextual embeddedness or has fewer associated cognitive operations. An individual who has not constructed the lie, but merely reports a ready-made one, is likely to require less cognitive effort to lie than an individual who constructs the lie themselves, as they do not need to engage cognitive processes for the retrieval, manipulation and organization of counterfactual information and can instead focus on lie delivery. The memory for an experimenter-generated lie is therefore likely to have fewer associated cognitive operations than the memory for a self-generated lie, thus increasing commission errors. Consistent with this interpretation, other research has shown that factors that are known to influence source monitoring ability affect the likelihood that lying will impair memory. One such factor is whether the lie is told merely once or repeated.

There is conflicting evidence regarding the effect of repeating fabrications on memory, depending on how source monitoring is affected. One study showed that repeating a fabrication twice increased participants' belief that the fabricated event occurred (Polage, 2018). This is consistent with research showing that repeating lies reduces the cognitive cost of lying (Van Bockstaele et al., 2012) and that lying can actually become easier than truth telling with practice (Hu, Chen, & Fu, 2012). The first time that an individual fabricates information, they must generate the false details to report, whereas repeating that fabrication on subsequent occasions instead requires that they simply retrieve what was previously said. Repeated retellings may enhance memory for the contents of the lie, while memory for the cognitive operations used to generate it might fade. Indeed, related research into the effects of imagination on memory shows that repeatedly imagining an event that did not occur increases belief that the event did indeed occur compared to when the event is imagined only once (Goff & Roediger, 1998). Thus, fabricated information might more closely resemble a genuine memory after it is repeated.

It should be noted, however, that participants in Polage (2018) repeated their fabrications just twice. Other research suggests that when lies are repeated more than twice, repetition may instead help protect against memory distortion, rather than facilitate it. In one experiment examining the effect of repeated lying on memory, participants watched a video and were interviewed about its contents immediately afterwards and again two days later. During interview, participants responded truthfully to questions concerning details that were present in the video, but had to fabricate answers to questions concerning details that were not present in the video. To distinguish this from a forced fabrication paradigm, participants were explicitly told that they would be asked about details that they had not seen and that they should fabricate answers to these questions. Some questions were asked once, whereas others were asked thrice so that participants repeated their lies for a subset of questions. Importantly, some of the items were not asked about at interview to determine baseline memory performance (Rindal, 2017, Exp 1).

In a recognition test taken two days later, participants were less likely to confuse their fabrication for the truth when they repeated those fabrications compared to when they told them only once, but still showed increased commission errors for repeated fabrications compared to control items. Thus, repeating fabrications still increased commission errors, but the most commission errors arose from single lies: Over four experiments, fabricating tripled commission errors compared to no rehearsal; participants confused their lie for the truth 30% of the time compared to a base rate of 10% (Rindal, 2017).

The finding that repeating lies can enhance source monitoring has been shown in another study that compared the effects of denial and fabrication on memory. Participants studied a series of images consisting of an object together with its label. They were then shown a subset of the object labels they had studied, as well as some new object labels, and either lied or told the truth about whether they had studied the corresponding object in the previous phase. Participants either truthfully or falsely denied seeing something that they had in fact studied or described something they had or had not studied. Additionally, some object labels were shown just once, whereas others were shown thrice so that participants repeated their truthful or deceptive answers for some of the objects.

Two days later, participants completed a memory test for all of the objects initially studied assessing both item memory (i.e., whether they correctly remembered studying the object), and source memory (i.e., whether they remembered talking about the object subsequently). Participants showed more accurate source memory when they described an object, rather than denied it. Source memory was further enhanced when subjects repeated their answer three times compared to when they answered just once (Vieira & Lane, 2013).

Subsequent research suggests that it may not be repetition *per se* that helps to protect against memory distortion, but instead that repetition enhances participants' memories for generating the fabrication (Lane, Dianiska, & Cash, as cited in Dianiska, Cash, Lane, and Meissner (2019)). Each time the lie is repeated it is likely that memory is reinforced for the act of lying, which serves as a reminder that the lie itself is indeed false information. Consistent with this interpretation, one study showed that the effect of fabrication on memory depended on participants' memory for having lied. Participants who fabricated a childhood event that never happened decreased their belief that they had experienced that event when their memory was tested one week later, but only when they remembered having lied about that event. Participants who forgot that they lied instead increased their belief that the event occurred, a finding called *fabrication inflation* (Polage, 2004). Interestingly, while most participants showed small changes in their belief ratings that the fabricated event occurred, a significant proportion of participants were described as "big jumpers": Over 2 experiments, 10% and 16% of participants shifted from a low belief rating that the event had happened before fabricating to a high belief rating after fabricating. This finding suggests that there may be individual differences in people's susceptibility to memory distortion following fabrication (Polage, 2004).

One individual difference that seems to affect the likelihood that fabrication will distort memory is source monitoring ability. In a second study investigating potential moderators of the fabrication inflation effect, participants who had a tendency towards dissociative experiences were more likely to increase their belief that the fabricated event occurred. Importantly, this finding was unique to fabricated events; high dissociators did not increase belief in childhood events that they had not lied about. A propensity towards dissociative experiences has been linked to poorer source monitoring ability (Hekkanen & McEvoy, 2002). Thus, the increased fabricated inflation in high-dissociators suggests that individuals who have difficulty source monitoring are more prone to memory distortion following fabrication. Participants who reported lying more often were also more likely to show fabrication inflation, which may be because they are likely to find lying easier and therefore may have fewer cognitive operations associated with their memory of the lie (Polage, 2012).

In sum, there is a growing body of evidence that fabricating counterfactual information can indeed promote commission errors—or false memories—and therefore that the “self-induced false memories” described by von Hippel and Trivers (2011) are possible. Moreover, the likelihood that fabricating will increase commission errors is affected by factors known to influence source monitoring. Specifically, people are more likely to mistake their lies for the truth when those lies are plausible (Pickel, 2004), not self-generated (e.g., Pezdek et al., 2009; Pickel, 2004), told on a single occasion (e.g., Rindal, 2017), and when the individual lies often or has a propensity towards dissociative experiences (Polage, 2012).

One important factor that is known to affect source monitoring in the broader memory literature is cognitive load. Specifically, participants who experience greater cognitive load during encoding or retrieval are more prone to memory errors (Knott & Dewhurst, 2007; Zaragoza & Lane, 1998). The relationship between cognitive load and deception-induced memory distortion has not yet been investigated. However, this relationship warrants special consideration, as the concept of cognitive load in deception is critical to the only existing theory that is dedicated to explaining the effect of lying on memory: The memory and deception (MAD) framework (Otgaar & Baker, 2018).

1.6 The memory and deception (MAD) framework

The MAD framework (Otgaar & Baker, 2018) is currently the only theory that is dedicated to explaining when and how lying can affect memory for the truth. Like von Hippel and Trivers' theory of self-deception, the MAD framework focuses on the relationship between deception and cognitive load, however it takes the opposite perspective on this relationship. Whereas von Hippel and Trivers see memory distortion as a strategy for reducing the cognitive load of deception, the MAD framework proposes that the high cognitive load of deception may in fact cause memory distortion.

The MAD framework focuses on three types of lie: [1] false denials (denying aspects of or the entirety of an event), [2] feigned amnesia (pretending to have forgotten about aspects of, or the entirety of, an event), and [3] forced fabrication (when an individual is compelled to provide false details, due to intimidation or an interviewer who forces the individual to guess unknown details). The framework is built on the observation that these different types of lie appear to affect memory in different ways. As outlined in Section 1.4, numerous studies show that when people falsely deny or feign amnesia, they are more likely to omit information in later truthful reports than they otherwise would if they told the truth all along (they make more omission errors relative to truth tellers; e.g., Van Oorsouw & Merckelbach, 2004; Vieira & Lane, 2013). For instance, an eyewitness who falsely denies that they saw the perpetrator or claimed to have forgotten what they looked like, might subsequently remember fewer details about their appearance than if they had truthfully reported them at the outset. In contrast, other studies show that when people are forced to fabricate alternative information, they often incorporate this information into memory later on (they make more commission errors; e.g., Chrobak & Zaragoza, 2009; Polage, 2012). To take the previous example, an eyewitness who fabricates that the perpetrator's hair was blonde when in fact it was brown might later falsely remember that their hair was blonde.

Why do false denials and feigned amnesia increase omission errors, but forced fabrications increase commission errors? The MAD framework explains this pattern of findings with reference to the cognitive resources required to lie. The central claim of the MAD framework is that lies requiring few cognitive resources increase the likelihood of omission errors, whereas lies requiring greater cognitive resources increase the likelihood of commission errors.

According to the MAD framework, both false denials and feigned amnesia require relatively few cognitive resources; individuals can simply respond "no" or "I don't remember" to any question. This type of response prevents the rehearsal of truthful details, reducing the likelihood that they will be recalled later on (e.g. Van Oorsouw & Merckelbach, 2004; Vieira & Lane, 2013). In contrast, fabrication requires more cognitive resources because the individual must conjure up counterfactual information and construct a story that is plausible and consistent. Like von Hippel and Trivers (2011), the MAD framework appeals to the SMF to explain how fabrication can lead to commission errors. By fabricating information, the individual creates a source monitoring problem, that is, they must later identify what information in memory was fabricated and what was truly experienced. If the fabricated information shares the subjective characteristics of a genuine memory, it may be mistaken for a memory of a real experience, leading to a commission error (e.g. Pickel 2004; Polage, 2012).

In sum, the MAD framework aims to predict the type of memory error from the cognitive effort required to lie. If the lie requires relatively few cognitive resources (e.g. false denial and feigned amnesia), we should expect an increase in omission errors, whereas if the lie requires more cognitive resources (e.g. forced fabrication), we should expect an increase in commission errors. There is currently no research that explicitly tests the hypotheses borne out of the MAD framework. Chapters 3 and 4 are dedicated to exploring this framework further and begin to test its hypotheses.

1.7 Conclusion

We know from a wealth of research that misinformation can distort memory. Although lying is a category of misinformation, the additional cognitive load generated by lying compared to other types of misinformation may lead to unique effects on memory. The effect of lying on memory therefore deserves independent consideration to other types of misinformation. While research in this area is gradually gaining traction, there is still much to be investigated.

The concept of cognitive load plays an important explanatory role in our understanding of how lying might affect memory and is a central concept to both of the theories in this area. For von Hippel and Trivers (2011), the central idea is that individuals will do what they can to reduce the cognitive cost of lying to elude detection. Self-deception may achieve this goal by reducing the accessibility of the truth via RIF, or by distorting memory for the truth via failures in source monitoring. These mechanisms can reduce the liar's cognitive load by alleviating the need to suppress the truth in memory or monitor their behavior, and therefore may help the liar to elude detection. Consistent with this idea, there is preliminary evidence that self-deception does indeed decrease liars' cognitive load (Jian, Zhang, Tian, Fan, & Zhong, 2019).

The MAD framework instead considers cognitive load from the opposite perspective by suggesting that liars' increased cognitive load might actually drive the memory impairment. While the concept of cognitive load forms the backbone of both theories, it has received no empirical attention in the deception and memory literature. The experiments presented within this thesis aim to test some of the claims of von Hippel and Trivers' theory of self-deception and the MAD framework to build the empirical literature in this area and contribute to our theoretical understanding of the relationship between lying and memory.

1.8 Research outline

As outlined in the above review, the effect of lying on memory has received relatively little attention from researchers in comparison to other types of misinformation. This is particularly true for deliberate fabrication– that is, when an individual knowingly communicates false details with the intention of deceiving another– for which there are just a handful of studies looking at its effect on memory. In addition, there is just one theoretical paper that focuses specifically on developing a theory for how lying can affect memory—the MAD framework.

Given the prevalence of deception within the criminal justice system, it is vital that a) investigators have the best chance of detecting deception where possible and are therefore aware of the factors that might influence accurate deception detection, such as memory distortion, and b) that accurate information remains retrievable from people who wish to retract an initially deceptive statement and respond honestly.

In light of the above, the research presented in this thesis focuses on deliberate fabrication in a forensic context. The current experiments investigate the effect of fabricating false information on memory for the truth, while also testing some of the purported mechanisms behind these effects to advance our theoretical understanding of the relationship between lying and memory.

In Chapter 2, Experiments 1-3 address the first research question of the thesis: Can fabricating false information promote RIF of the corresponding truthful information? These experiments are the first in the field that are specifically designed to test for an inhibitory mechanism that may promote forgetting of the truth when people fabricate information.

Chapter 3 begins to explore the MAD framework by challenging its proposal that the cognitive demand of lying may predict the memory errors that can follow. This chapter proposes that centering the framework on cognitive load could lead to misleading interpretations of the empirical findings.

In Chapter 4, Experiments 4 and 5 test the claims made in Chapter 3. Here, the second and third research questions of the thesis are addressed: Do cognitively demanding lies increase commission errors and what is the relationship between the cognitive effort required to lie and source monitoring ability?

Chapter 5 reports the sixth experiment of the thesis to address the final research question of whether fabricating false information in an initial interview can hamper subsequent lie detection by impairing memory for the truth. This experiment investigates whether the accuracy of a popular lie detection test—the concealed information test (CIT)—can be reduced when an individual has fabricated false information in a previous interview.

Finally, I bring together the findings from the 6 experiments presented throughout the thesis in the general discussion in Chapter 6. Here, I discuss the implications of the findings for the two existing theories that hypothesize about the relationship between lying and memory—Von Hippel and Trivers' theory of self-deception and the MAD framework, as well as the potential practical considerations for deception detection.

Chapter 2: Can repeating fabrications promote retrieval-induced forgetting of the truth via an inhibitory mechanism?

2.1 Introduction

A good liar is one who practices. There are several reasons for this: An individual who practices their lies can embellish them with details in advance, respond to questions faster and more fluently, and is more likely to remain consistent over time. Since detail, speed and consistency are associated with honesty, practicing lies increases credibility, ultimately reducing the likelihood of detection (DePaulo et al., 2003; Granhag & Strömwall, 2000; Walczyk, Mahoney, Doverspike, & Griffith-Ross, 2009). A good liar, then, strengthens their memory for the lie so that it is readily available given the right cue, for instance, an investigator's question. But what effect does repeating a lie have on the accessibility of the truth? While several researchers have considered that repeating lies might render the truth less accessible in memory via the process of RIF (e.g., Christianson & Bylin, 1999; Gronau et al., 2015; Paige et al., 2018; Polage, 2018), this proposition remains untested. We now know that the cognitive processes that promote RIF are also typically engaged when people fabricate false information. Thus, it is plausible that repeating fabrications might indeed reduce the accessibility of the truth via RIF. Research into the cognition of deception can explain how this might occur. An overview of this research is given below.

2.1.1 The activation-competition-inhibition cycle of RIF

As detailed in Chapter 1, RIF is an extensively investigated and robust memory phenomenon (Murayama et al., 2014). Since RIF is defined and explained in Section 1.4, only a brief overview of the phenomenon is given here, and this section instead focuses on the specific processes that lead to RIF. So to recap, RIF is the finding that practicing retrieval of previously learnt information enhances memory for that information, but impairs memory for related information that was not practiced (Anderson et al., 1994). The forgetting is "retrieval-induced" because it is caused by practicing retrieval of other information. As outlined in Section 1.4, it is only inhibitory based RIF that represents true forgetting that generalizes across different contexts and modes of testing, whereas competition-based RIF represents accessibility issues that are specific to a given test (Anderson, 2003; Chan et al., 2015). The experiments in this Chapter therefore focus exclusively on the question of whether repeating fabrications can promote RIF of the truth via an inhibitory mechanism.

According to the inhibitory account of RIF (Anderson, 2003), there is a 3-stage process that causes forgetting, which I term the *activation-competition-inhibition cycle*:

[1] Activation: Attempting to retrieve a target item in memory activates other items in memory that share a retrieval cue, but are not the targets for recall
[2] Competition: The activated non-targets compete with the target for retrieval and therefore interfere with target retrieval
[3] Inhibition: Activated non-targets are inhibited to resolve the interference and facilitate target retrieval, which weakens memory for the inhibited items

Note that the initiating component for RIF is that the target shares a retrieval cue with other items in memory. Thus, when the individual is presented with the cue, multiple items are activated, inducing the subsequent processes that ultimately lead to forgetting. To recall the example from Section 1.4, it is because the items "orange" and "lemon" share the retrieval cue "fruit" that practicing retrieval of "orange" impairs memory for unpracticed items like "lemon", but not unpracticed items like "whisky".

Fabricating false information shares this initiating component of RIF, since it also generates a situation where there are multiple items in memory associated with the same retrieval cue. To illustrate, if an investigator asks a suspect what time they arrived home on the night of the crime, this question will likely cue both the truth (since it directly solicits it) and a fabricated alternative (since the suspect does not want the investigator to know the truth). If the suspect is asked this question on multiple occasions and responds with the same fabrication each time, the same conditions that lead to RIF are present: There are multiple items in memory that are associated with the same retrieval cue, but only one item is selectively practiced (the lie), while the related information remains unpracticed (the truth). The question is whether there is a truly analogous situation to RIF, whereby selective practice of the lie impairs later recall of other associated items in memory (i.e., the truth). There are numerous reasons to think that this might be so. Research into the cognition of deception suggests that the same activation-competition-inhibition cycle is also present when people fabricate. In the context of fabrication, the target response is the lie and the non-target is the truthful item in memory.

2.1.2 Activation, competition, and inhibition in deception

A substantial body of research demonstrates that the truth is automatically activated in the initial stages of deception and that it plays an important role in lying successfully. While some might consider the truth an inconvenience, the best liars use it to their advantage. Liars frequently report staying close to the truth as a strategy for avoiding detection (Leins, Fisher, & Ross, 2013; Vrij, Granhag, & Mann, 2010). People also use the truth to aid lie construction, either to cue related plausible information or to respond with the opposite answer (Walczyk et al., 2014; Walczyk, Roper, Seemann, & Humphrey, 2003). Constructing plausible lies therefore requires that the truth is retrieved and active in working memory (Christ, Van Essen, Watson, Brubaker, & McDermott, 2009; Debey, De Houwer, & Verschuere, 2014; Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, & Crombez, 2017; Walczyk et al., 2014; Williams, Bott, Patrick, & Lewis, 2013). Since lying requires withholding the truth, some researchers claim that lying *depends* on accurate retrieval of the truth and that the deceiver must be conscious of the truth to generate a deceptive response (Johnson, Barnhardt, & Zhu, 2004).

Consistent with this reasoning, numerous studies show that the truthful (non-target) response is the dominant response that is activated on receiving a retrieval cue, such as a question. For instance, in one study, participants responded either truthfully or deceptively to yes/no autobiographical questions by steering a Nintendo Wii remote to either side of a screen. Analysis of arm movement coordinates revealed that participants were initially drawn towards the truthful response when answering deceptively. Importantly, they did not tend towards the opposite answer when responding truthfully, but rather moved directly towards the true answer (Duran et al., 2010). This suggests that lying requires the individual to overcome their automatic tendency to respond truthfully. There are also consistently higher error rates in lie trials compared to truth trials in reaction-time-based deception studies (Debey, Ridderinkhof, De Houwer, De Schryver, & Verschuere, 2015; Johnson et al., 2004; Suchotzki, Crombez, Smulders, Meijer, & Verschuere, 2015) This suggests the presence of a dominant truthful response that conflicts with the goal to respond deceptively, leading to more errors.

Despite its importance for constructing plausible lies, automatic activation of the truth can create competition because the deceptive response is the recall target, not the truth. The response conflict that arises from this competition is evident in the brain scans of deceivers. Imaging studies show that deceptive responding recruits brain regions such as the anterior cingulate cortex and dorsolateral prefrontal cortex, which are associated with conflict detection, cognitive control and response selection (Abe, 2011; Sun, Mai, Liu, Liu, & Luo, 2011; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001). Lying successfully therefore requires a mechanism to overcome response conflict and enable execution of an untruthful response. As in RIF, research indicates that this mechanism is inhibition.

Evidence for the role of response inhibition in deception primarily comes from imaging and ERP studies. The ventromedial prefrontal cortex is another brain region consistently implicated in deceptive responding (Christ et al., 2009; Spence et al., 2001) and we know that lesions to this area result in difficulty inhibiting prepotent responses (Iversen & Mishkin, 1970). Similarly, ERP signatures of deceptive responses overlap with those of classic response inhibition tasks, such as the Go No-Go and Stroop tasks (Johnson et al., 2004; Suchotzki et al., 2015). Other studies using techniques that more directly measure inhibitory processes, such as delta plot analysis, also demonstrate an important role for inhibition during deception (Debey et al., 2015; Williams et al., 2013).

Taken together, research into the cognition of deception suggests that the cognitive processes recruited when people fabricate mirror those underlying RIF and therefore that repeating fabrications could promote inhibition of the truth, thereby reducing its accessibility in memory. Although previous research has speculated that RIF could underlie memory impairments following deception, no research has explicitly tested this possibility using a procedure that is designed to elicit RIF. In the experiments that follow, the basic retrieval practice paradigm is modified to investigate if repeating fabrications can impair memory via RIF.

2.2 Experiment 1a

The basic retrieval practice paradigm consists of 3 main phases: Learning, retrieval practice, and a recall/recognition test (Anderson et al., 1994). In the learning phase, participants study category-item pairs, such as “Fruit-Orange”, “Fruit-Lemon”, “Drink-Whisky”, etc. Participants then complete the retrieval practice phase, where they are prompted to repeatedly retrieve half of the items from half of the categories by completing category-stem pairs (e.g., “Fruit-O__”). After a short delay, participants’ memories are tested for all category-item pairs initially studied.

This procedure creates three types of item: Items practiced in the retrieval practice phase (Rp+ items, e.g., “Orange”), items not practiced in the retrieval practice phase, but belonging to the same category as those that were practiced (Rp- items, e.g., “Lemon”), and items belonging to categories that were not prompted at all in the retrieval practice phase (Nrp items, e.g., “Whisky”). RIF refers specifically to the finding that memory for Rp- items (unpracticed items related to those practiced) is significantly poorer than memory for Nrp items (which represents baseline memory).

There are two critical components of RIF studies that are required to conclude that an inhibitory mechanism underlies memory impairment: First, given that RIF refers specifically to below-baseline forgetting, the experiment must measure memory for information that is both unpracticed and unrelated to practiced information to determine participants’ baseline forgetting rate. In the present experiment, if memory performance for the truth falls below this baseline measure after participants practice their fabrications, it suggests the existence of an inhibitory mechanism that accelerates forgetting. If instead memory performance does not differ from baseline, it suggests that any forgetting is merely due to the passing of time and that fabricating does not have any unique effects on memory. Second, the procedure must control for output interference in the final memory test so that any difference in recall performance can be attributed to changes in memory, as opposed to the order in which information is retrieved at test (the importance of controlling for output interference is fully explained in Experiment 1b).

To meet the above conditions and investigate if repeating fabrications can impair memory via an inhibitory mechanism, the basic retrieval practice paradigm was modified to include a fabrication condition. The paradigm was kept as close to the original as possible (as in Anderson et al., 1994) to maintain a high level of experimental control and enable a straightforward comparison with previous RIF research. As such, participants learnt category-item pairs (e.g., "Clothing-Shorts", "Sport-Tennis") and subsequently practiced retrieving a subset of those pairs either truthfully or deceptively by completing word stems (e.g., "Clothing-S_ _"). Deceptive retrieval involved fabricating an alternative category member beginning with the same letter (e.g., "Skirt") and repeating this fabrication each time it was prompted. After a short delay, participants completed a surprise final recall test assessing their memory for all items initially learnt. If repeating fabrications promotes inhibition of the corresponding truthful information, this should lead to poorer memory for items participants practiced fabricating about compared to baseline memory.

2.3 Method

2.3.1 Participants.

The experiment was completed online by 151 participants in exchange for entry to a prize draw. Data from 17 participants were excluded: 13 participants failed to complete the experiment and 4 participants failed to correctly recall a single item in the retrieval phase, indicating that they had not encoded the material that they were instructed to learn. This left data from 134 participants for analysis (81 female, 51 male, 2 preferred not to say; 106 aged 18-25, 12 aged 26-34, 11 aged 35-54, 3 aged 65+ and 2 preferred not to say).

Participants were asked to rate their fluency in English to confirm that they were fluent to an excellent or native standard. If participants selected an option lower than "excellent", the experiment was aborted (104 and 30 participants rated themselves as fluent or excellent in English respectively). The study was approved by the Psychology Department ethics committee at the University of Warwick.

2.3.2 Stimuli.

The category-item pairs were selected from published word norms (Van Overschelde, Rawson, & Dunlosky, 2004). Categories were selected using similar constraints as Anderson et al. (1994): They were a single word in length and semantically and phonetically unrelated. The items in each category were selected to have a mid-taxonomic strength (ranked between 4-7 in Van Overschelde et al.), so that the category-stem cue would successfully cue the item, without the association between the category and its item being so strong that interference from other items would never occur. The items within each category began with different letters so that no two items would have the same category-stem cue in the retrieval practice or recall phases.

A pilot questionnaire ensured that participants (n=18) could consistently generate a lie for each item; that is, at least one alternative category member beginning with the same letter as the item studied (see Appendix 1 for the results). Based on the pilot, 6 items were selected from 6 categories, yielding 36 critical items. A further 3 items from 2 additional categories were selected to yield 6 filler items. Participants therefore learnt a total of 42 category-item pairs (see Appendix 2).

2.3.3 Design and procedure.

An overview of the procedure is depicted in Figure 2.1. The experiment was created using Qualtrics software (Qualtrics, Provo, UT) and conducted online. The experiment had a 2 (Veracity: truth, lie) x 3 (Retrieval Practice Status: Rp+ [practiced], Rp- [unpracticed, related to practiced], Nrp [unpracticed, unrelated to practiced]) within-subjects design. The experiment consisted of 4 phases: Learning, retrieval practice, distractor and surprise final recall. Participants were informed that they were participating in an experiment on memory and reasoning in deception.

Learning. Each of the 42 category-item pairs was presented one-by-one in the center of the screen. Participants were instructed to learn each pair and informed that their memory for the pairs will be tested. The first and last 3 pairs were fillers to control for primacy and recency effects. Critical category-item pairs were presented in a blocked semi-random order. There were 6 blocks each consisting of one item per category, yielding 6 category-item pairs per block. Each pair was on screen for 4 seconds and there was a 5s break between blocks. The order of the pairs within each block was randomized and the block order was counterbalanced such that half of the participants initially learned blocks 1-3 and the remaining half initially learned blocks 4-6. Once all blocks had been presented, each block was repeated, once again with the order of items within each block randomized and the block order counterbalanced. Participants therefore saw each category-item pair twice for 4s each time, yielding a total of 8s learning time per pair.

Retrieval practice. Instructions were provided for the retrieval practice phase and participants were asked 3 questions as a comprehension check. Participants could only proceed if they answered all 3 questions correctly. In this phase, category-stem cues (e.g., "Sport - T__") were presented one-by-one in the center of the screen for a subset of the items learnt (described in more detail below). The task was to complete each stem with the truthful answer—the item previously learnt—or a deceptive answer—an alternative category member beginning with the same letter that participants fabricated themselves. An instruction (LIE/TRUTH) at the top of the screen informed participants to answer truthfully or deceptively for each trial.

Participants first completed 3 filler category-stem cues as practice trials. They then truthfully completed stems for half of the items from 2 critical categories and deceptively completed stems for half of the items from 2 critical categories (and so they were prompted to retrieve a total of 12 different items from 4 categories). Each of the 12 category-stem cues was presented 4 times so that participants practiced retrieving either the truthful or deceptive response for each pair. For lie trials, participants were instructed to provide the same fabrication each time it was prompted.

As in Anderson et al. (1994) the order of category-stem cues was fixed to an expanding schedule, with an average of 4.8 trials between the first and second test, 10.4 trials between the second and third test and 11.3 trials between the third and fourth test. Trials were also organized so that no two categories were tested consecutively. Participants were given a maximum of 1-minute per cue to allow enough time to fabricate alternatives in lie trials. If no response was registered within 1-minute, the program moved on to the next category-stem cue. The specific category-stem cues presented were fully counterbalanced across participants so that every category-item pair served in each condition.

This procedure generates 5 types of item:

- Rp+ Truth – items participants practiced truthfully
- Rp+ Lie – items participants practiced lying about
- Rp- Truth – items participants did not practice, but were in the same category as truthfully practiced items
- Rp- Lie – items participants did not practice, but were in the same category as items lied about
- Nrp – items participants did not practice belonging to categories that were not cued at all during retrieval practice

Distractor. Participants completed a variety of mathematical and reasoning tasks (e.g., sudoku, counting the number of triangles in a picture) for 10-minutes, then the experiment automatically proceeded to the surprise final recall phase.

Surprise final recall. Category-stem cues (e.g., “Sport-T__”) for all 36 critical items were presented one-by-one in the center of the screen. The task was to complete each stem with the truthful answer, that is, the item originally learnt. A comprehension check ensured that participants understood this before they could proceed. The stems were presented in a random order across participants. This aspect of the procedure is important: Randomizing the order of the category stem cues means that participants cannot choose the order in which they recall the items, which controls for output interference. This means that we can be more confident that any forgetting effects are due to inhibition of items in memory during retrieval practice, as opposed to interference caused by retrieving the strongest items in memory first when participants can choose the order of recall (Murayama et al., 2014).

After completing each stem, participants rated their confidence that their answer was the item they had originally learnt on a sliding scale of 0-100, where 0 = “completely uncertain” and 100 = “completely certain”. Participants completed the task at their own pace, but were given a maximum of 1 minute to complete each stem and rate their confidence. The experiment automatically progressed onto the next stem if no response was registered within one minute.

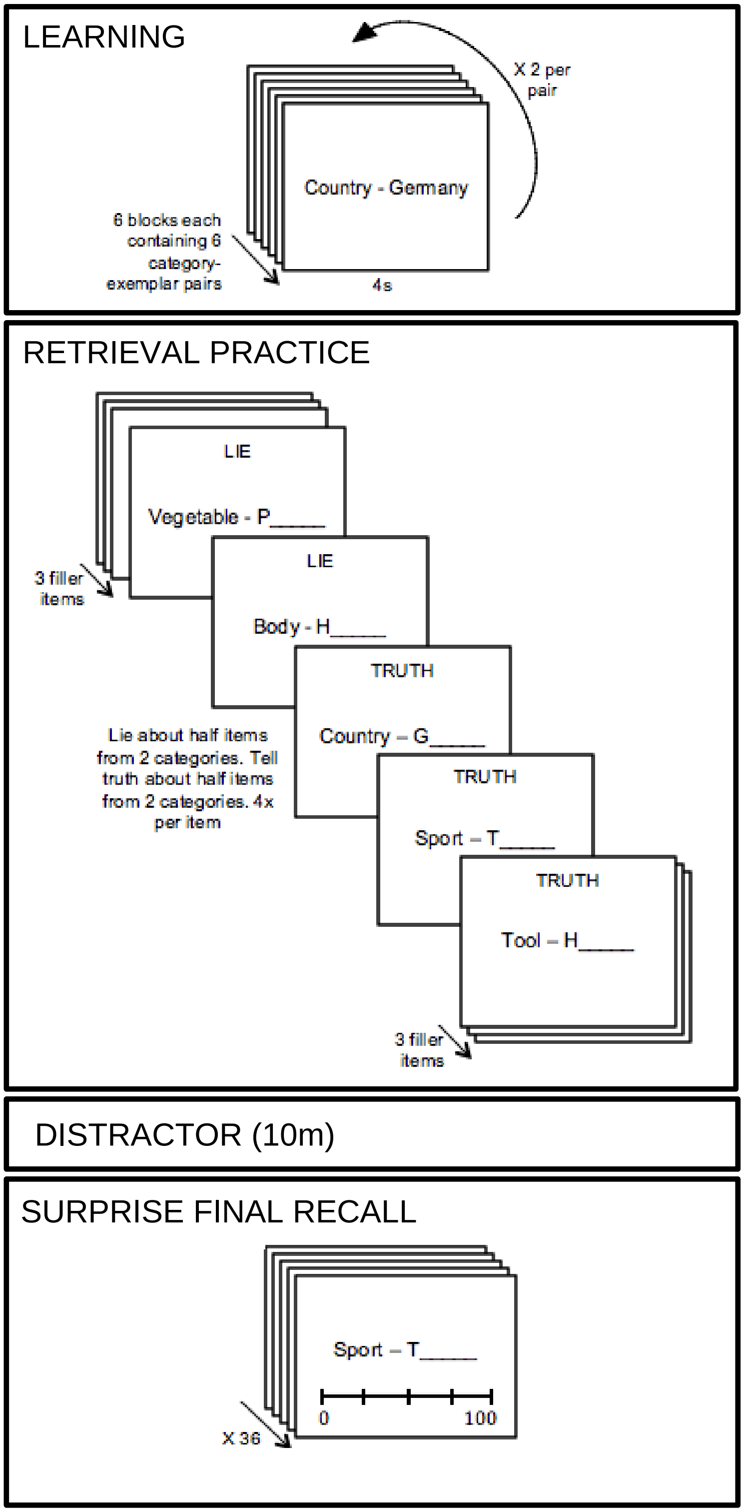


Figure 2.1. Procedural overview of Experiment 1a.

2.4 Results

2.4.1 Retrieval practice performance.

Performance in truth trials of the retrieval practice phase indicates how well participants encoded the category-item pairs in the learning phase. Participants performed moderately well in truth trials, correctly answering an average of 68.12% (*SD* = 24.14%) of trials. However, this indicates that participants failed to encode almost a third of the items in the learning phase. Performance in the lie trials indicates that participants generally adhered to the lie instruction by providing a correct alternative category member: Participants correctly answered an average of 88.74% of lie trials (*SD* = 18.62%).

In the retrieval practice phase, participants were prompted to retrieve each item 4 times. To include the recall data for an item in the final analysis, participants had to answer at least 3 out of four trials for that item correctly. Final recall data for items that did not meet this criterion were excluded from analyses. Three was chosen as a cut-off point to minimize data loss while also only including items for which participants had repeated their lies. This led to the removal of 11.68% of the final recall data for liars and 32.13% for truth tellers.

2.4.2 Final recall performance.

A 2x3[1] repeated-measures ANOVA on the proportion of items correctly recalled revealed a significant main effect of Veracity ($F(1, 256.52$[2]$) = 85.39$, $p < .001$) and Retrieval Practice Status ($F(1.96, 256.52) = 73.03$, $p < .001$). Main effects were superseded by a significant interaction, thus the effect of retrieval practice on the proportion of items correctly recalled depended on whether participants lied or told the

[1] Since items in the Nrp condition were not included in the retrieval practice phase (by definition), participants neither lied nor told the truth about these items. Nrp items therefore do not have a veracity status, i.e., there are no lie trials and truth trials for Nrp items. As the design was fully within-participants, this resulted in a fractional factorial design consisting of 5 cells: Rp+ Lie, Rp+ Truth, Rp- Lie, Rp+ Truth, and Nrp. To enable a 2x3 ANOVA to be performed, the first step of analysis was therefore to make the design fully factorial by splitting the Nrp trials into two levels. Each Nrp trial was randomly designated a lie or truth trial to create two levels (NrpLie and NrpTruth). A paired t-test indicated that there was no significant difference in the proportion of items correctly recalled between Nrp trials that were randomly allocated to truth/lie levels of the Veracity variable ($t(133) = -0.22$, $p = .83$).

[2] Huynh-Feldt-corrected degrees of freedom are reported due to violation of the assumption of sphericity: $\chi2(2) = 6.99$, $p = .03$, $\varepsilon = .96$).

truth (F(1.93, 256.52) = 60.74, $p < .001$). To break down this interaction, the data were split on the Veracity variable and simple main effects analyses were performed for lie and truth trials separately. The findings are reported below.

Recall performance in lie trials. The hypothesis of this experiment was that repeating fabrications would promote inhibition of the corresponding truthful item and that this would impair later recall for that truthful item. For lie trials, Rp+ represents memory for the items that participants repeatedly fabricated about. Support for the hypothesis therefore requires significantly poorer memory for Rp+ items in lie trials compared to baseline memory (Nrp). However, we can see from the top row of Table 2.1 that participants did not recall fewer items as a result of repeatedly fabricating about them. Indeed, there was no significant effect of retrieval practice status on the proportion of items correctly recalled (F(2, 266) = 2.70, $p = .07$). Repeating fabrications did not affect memory for Rp+Lie items (compared to NrpLie items; $p = 1$, 95% CI [-.05, .08]).

We can see from Table 2.1 that participants recalled fewer Rp-Lie items compared to baseline. Rp- refers to items that participants were not questioned on, but are related to lied-about items, as they belong to the same category. This impairment was, however, not significant ($p = .29$, 95% CI [-.10, .02]). Thus, practicing lies did not affect recall performance.

Recall performance in truth trials. Performance in truth trials was analyzed to determine a typical RIF effect was elicited; that is, lower recall performance for Rp- items compared to baseline memory (Nrp items) in truth trials. The proportion of items correctly recalled was affected by the retrieval practice status of the item (F(1.94, 257.40) = 128.88, $p < .001$; Huynh-Feldt-corrected degrees of freedom are reported due to violation of the assumption of sphericity $\chi^2(2) = 6.50$, $p = .04$, $\varepsilon = .97$).

Participants correctly recalled significantly more items after truthfully practicing them, as indicated by a Bonferroni-adjusted pairwise comparison of Nrp and Rp+ items ($p < .001$, 95% CI[.29, .42], $d = 1.51$). This simply reflects a rehearsal effect (Karpicke & Roediger, 2008). However, we can see from the second row of Table 2.1 that the typical

RIF effect was not replicated: Recall performance was not significantly lower for Rp- compared to Nrp items (p = 1, 85% CI[-.06, .07]).

Bayesian analysis. Bayesian analyses were performed to determine the likelihood that the null hypothesis obtains (that is, that practicing lies does not impair memory compared to baseline), as opposed to a Type II error. Bayesian analysis quantifies the likelihood that one hypothesis is true over another by first assuming a prior probability that one hypothesis holds and then updating this probability from modelling the observed data. The resulting Bayes Factor tells us how much more or less likely it is that the observed data would be obtained if the null hypothesis were true (BF_{01}) or if the alternative hypothesis were true (BF_{10}; van Doorn et al., 2019). A Bayes Factor of 1 indicates equal support for the null and alternative hypotheses. A BF_{01} of 1-3 typically represents anecdotal evidence in favor of the null, a BF_{01} of 3-10 represents substantial evidence in favor of the null, a BF_{01} of 10-30 represents strong evidence in favor of the null and anything above 100 is considered decisive evidence for the null (Jarosz & Wiley, 2014). All Bayesian analyses were conducted using Jasp v0.9.1.0 (Jasp Team, 2019).

Bayesian analysis for lie trials. To determine the likelihood that the null hypothesis obtains, a one-way Bayesian repeated-measures ANOVA was performed on participants recall performance in Rp+, Rp- and Nrp trials for items that participants lied about. The prior was set to the default used by Jasp software (Jasp Team, 2019). This revealed a BF_{01} = 3.0 in favor of the null hypothesis, indicating that the data obtained are 3 times more likely to have occurred under the null hypothesis than the alternative hypothesis. This therefore provides substantial evidence that repeatedly fabricating about an item does not affect subsequent recall performance. Post hoc tests with a corrected prior to adjust for multiple comparisons (as performed by JASP software [Jasp Team, 2019]) provided strong evidence that repeating fabrications does not affect memory for the items lied about (BF_{01}= 8.50 for comparing Rp+Lie and NrpLie items). However, the evidence that repeating fabrications does not affect memory for items related to those lied about was merely anecdotal (BF_{01} = 2.70 for comparing Rp-Lie and NrpLie). This suggests that the non-significant result for the poorer recall performance for Rp- items compared to NrpLie items may indeed be due to a Type II error.

Bayesian analysis for truth trials. A Bayesian analysis was performed to determine the likelihood that the failure to replicate the typical RIF effect is a Type II error. Because the only comparison of interest in truth trials is between Rp-Truth and NrpTruth items, a Bayesian t-test was performed with the default prior used by Jasp Software (Jasp Team, 2019). This revealed that the data obtained are more than 10 times more likely under the null hypothesis that there is no difference in memory performance for Rp- and Nrp items, compared to the alternative hypothesis (BF_{01} = 10.26). This provides strong evidence that the null finding was not a Type II error and therefore that truthfully rehearsing information is unlikely to promote forgetting of related unrehearsed information using this procedure (Jarosz & Wiley, 2014).

Table 2.1.

Mean proportion of items correctly recalled and mean confidence ratings (scale 0-100) in the final recall test in Experiment 1a as a function of Retrieval Practice Status and Veracity.

	Proportion recalled (SD)			Confidence (SD)		
	Nrp	Rp-	Rp+	Nrp	Rp-	Rp+
Lie trials	.60 (.27)	.56 (.27)	.62 (.27)	70.47 (22.69)	67.55 (23.37)	74.80 (20.80)
Truth trials	.61 (.30)	.61 (.25)	.96 (.13)	72.49 (23.23)	69.64 (23.40)	89.64 (15.47)

Note: "Rp+" represents practiced items (practiced lies or practiced truths), "Rp-" represents unpracticed items that are related to practiced items, and "Nrp" represents unpracticed items that are unrelated to practiced items and therefore serves as the baseline to which other groups are compared.

2.4.3 Confidence.

Participants' confidence in their memories was assessed to determine if practicing lies leads to more subtle changes in memory than impairing explicit recall. Confidence ratings for trials where participants did not provide an answer or answered "don't know" were excluded from analysis, as the analysis concerned participants' confidence that the item they recalled was correct. This led to the removal of 24.16% of the data (13.63% of lie trials and 10.53% of truth trials).

Table 2.1 shows participants' mean confidence ratings (0 = *completely uncertain* to 100 = *completely certain*). Confidence was generally high and minimally affected by the experimental condition. Nonetheless, the effect of retrieval practice on confidence was affected by whether participants lied or told the truth, as indicated by a significant interaction between Retrieval Practice Status and Veracity in a 2x3 repeated measures ANOVA ($F(2, 264) = 8.69$, $p < .001$). Again, simple main effects analyses were performed on lie and truth trials separately to break this interaction down and the findings are reported below.

Confidence ratings in lie trials. Confidence significantly differed across item types ($F(2, 264) = 8.69$, $p < .001$). Participants were significantly more confident in their memories for items they had practiced lying about compared to baseline, as indicated by Bonferroni-adjusted pairwise comparison of Rp+ and Nrp items ($p = .03$, 95% CI[0.24, 8.43], $d = 0.20$). This adds further evidence that practicing lies did not promote forgetting of the corresponding truthful items. Although participants were marginally less confident in their memories for items related to those they lied about (Rp- items) compared to baseline (Nrp items), this reduction was not significant ($p = .29$, 95% CI[-7.14, 1.30]).

Confidence ratings in truth trials. Confidence significantly differed across item types ($F(2, 266) = 70.77$, $p < .001$). Unsurprisingly, participants were most confident in their memories for items they practiced retrieving truthfully compared to baseline, as indicated by Bonferroni-adjusted pairwise comparison or Rp+ and Nrp items ($p < .001$, 95% CI[12.69, 21.57], $d = 0.87$). Once again, this simply reflects a rehearsal effect.

Confidence did not significantly differ for items related to those truthfully practiced (Rp-) compared to baseline (Nrp items). This follows the same pattern of results as the recall data. Thus, there was no evidence of a typical RIF effect in recall or confidence ratings.

2.5 Discussion

Repeating fabrications did not impair recall or decrease confidence in memory for the corresponding truthful item. Memory performance for items that participants repeatedly lied about did not differ from baseline, indicating that repeating fabrications was equivalent to no rehearsal in the time between learning and recall. Thus, using a procedure that is designed to elicit RIF, there was no evidence of an inhibitory mechanism that increases omission errors when people repeat their fabrications. RIF also did not occur in truth trials and the experiment therefore failed to replicate the typical RIF effect.

One possibility for why RIF was not found in neither lie nor truth trials is that the association between the categories and each item might not have been strong enough to elicit RIF. RIF results from inhibition that is recruited to resolve retrieval competition between items in memory, and so, for RIF to occur, the retrieval cue must activate both the target for recall and other associated non-targets to create the conditions that lead to forgetting (Rowland, Bates, & DeLosh, 2014; Storm, 2011).

Eliciting a typical RIF effect in truth trials therefore requires that the retrieval cue—the category-stem cue (e.g., “Sport-R__”)—activates both the target (e.g., “Running”), as well as other studied non-target category members (e.g., “Tennis”), leading to competition for recall and therefore inhibition of the non-targets. However, if the association between the category and the studied items is too weak, the cue will not activate multiple items in memory and the target is therefore retrieved without needing to engage inhibitory processes.

Eliciting RIF in lie trials instead requires that the retrieval cue—e.g., “Sport-R__”—activates both the target lie response (e.g., “Rugby”) and the associated non-target truth response “e.g., “Running”). It is only when both the lie and the truth are activated in memory that the truth should compete for recall and therefore require inhibiting. Thus, if the association between the category and the truthful item is too weak, the lie can be retrieved without any interference from the truth, obviating the need for inhibition.

Consistent with this idea, research has shown that RIF effects are larger when items are strongly associated with their category (Anderson et al., 1994; Storm, Bjork, & Bjork, 2007). Nonetheless, the category items selected had mid – not weak – taxonomic strength and past research shows that RIF still occurs with weaker items, but merely to a lesser extent (Storm et al., 2007). In addition, the within-subjects design of the present experiment meant that participants were switching between lying and truth telling, and past research suggests that this can affect the retrieval and inhibitory processes engaged when people lie (Van Bockstaele et al., 2012). As such, the failure to find forgetting in the present study might be due to the procedure, rather than the stimuli.

The modified paradigm used here differed from the original retrieval practice procedure in several ways. The most significant departure was the addition of a lie condition, but more minor departures include the number of repetitions and the length of the category-stem cue. Participants practiced retrieving the item – truthfully or deceptively – 4 times, whereas Anderson et al. (1994) included only 3 repetitions. Additionally, only the first letter of each item was provided as a cue, whereas Anderson et al. provided the first 2 letters. To identify whether the results of the present experiment were due to the stimuli or procedure, a second experiment was conducted using the original retrieval practice paradigm, but with the present stimuli. Thus, the deception condition was removed and the original retrieval practice procedure was used.

2.6 Experiment 1b

Experiment 1b consisted of a learning, retrieval practice, and distractor phase, followed by one of two recall tests: Participants completed either a category-stem-cued recall test or a category-only-cued recall test. The category-stem-cued recall test was identical to that used in Experiment 1a, except the stem consisted of two letters instead of one (e.g., “Sport-Ru__”). The category-only-cued recall test instead provided only the category name and participants listed all items they could recall from the learning phase.

Studies using category-only-cued recall find RIF more consistently, and with a larger effect size, than those using category-stem-cued recall (Murayama et al., 2014). This is because category-only-cued recall confounds inhibitory effects with output interference. When freely recalling category items, participants choose the order in which items are recalled. This means that they are more likely to recall practiced items first because they are most easily retrieved. However, this can block retrieval of unpracticed items, resulting in below baseline memory performance for such items. Controlling the order of recall by providing category-stem cues eliminates output interference so we can be more confident that any memory impairment is due to inhibition during retrieval practice, rather than output interference at final recall. Both types of recall test were included to determine if RIF can be elicited under any circumstances using the stimuli from Experiment 1a.

2.7 Method

2.7.1 Participants.

The experiment was completed online by 80 Amazon Mechanical Turk workers in exchange for a small monetary reward. One participant failed to complete the experiment and their data were excluded, leaving data from 79 participants for analysis (38 female, 40 male and 1 preferred not to answer; 11 aged 18-25, 35 aged 26-34, 25 aged 35-54, 6 aged 55-64, 1 aged 65+ and 1 preferred not to say). As in Experiment 1a, participants were able to proceed to the experiment only if they rated their fluency as “fluent” or “excellent”. Participants were not permitted to complete the experiment if they rated their fluency lower than excellent (73 were native English speakers, 1 rated themselves fluent to a native standard and 5 rated their fluency as excellent). The study was approved by the Psychology Department ethics committee at the University of Warwick.

2.7.2 Stimuli.

The stimuli were identical to those used in Experiment 1a (see Appendix 2).

2.7.3 Design and procedure.

The experiment had a 3 (Retrieval Practice Status: Rp+, Rp-, Nrp) x 2 (Recall Test: category-cued, category-stem-cued) mixed design. The recall test was tested between subjects. The dependent measure was the proportion of items correctly recalled in the final recall test. The experimental design and procedure was the same as Experiment 1a, except for the following changes to mimic the original RIF procedure in Anderson et al. (1994):

Learning. Each category-item pair was presented only once for 5 seconds.

Retrieval practice. Participants truthfully completed category-stem cues for half of the items from 3 of the 6 categories (9 practiced items in total). The word stems showed the first 2 letters of the item, rather than only the first letter (e.g., "Sport-Te__") and each category-stem pair was presented 3 times. If the stem was not completed in 10 seconds, the experiment moved onto the next item.

As in Experiment 1a, the category-stem cues were counterbalanced across participants so that every category-item pair served in each condition. All stems were completed truthfully. Participants were tested on an expanding schedule: there was an average of 3.1 trials between the first and second test of a given item and 6.2 trials between the second and third test (a similar spacing to Anderson et al. (1994), who used an average of 3.5 trials and 6.5 trials respectively). On completion, participants proceeded to the distractor phase, which was identical to Experiment 1a.

Final recall. Participants were randomly assigned to the category-cued (n=39) or category-stem-cued recall test (n=40). In category-cued recall, each category name was sequentially presented at the top of the screen (e.g., "Sport") and participants listed all items belonging to the category that they could recall from the learning phase. The experiment moved onto the next category after 30 seconds. The first category was always a filler category and the order of the subsequent critical categories was randomly determined across participants.

In category-stem-cued recall, category-stem cues (e.g., "Sport-Te___") were presented one-by-one in the center of the screen for each of the 36 critical items studied in the learning phase. This followed an identical format to the retrieval practice phase: The category cue consisted of the category name together with the first two letters of the item and participants had 10 seconds to complete each stem. The first two category-stem cues were filler items and the subsequent 36 critical item cues were presented in a random order across participants.

2.8 Results

2.8.1 Retrieval practice performance.

Participants performed moderately well in the retrieval practice phase, answering correctly in an average of 70.89% of trials (*SD* = 20.93%). However, this still leaves a significant proportion of items that were incorrectly recalled, indicating that participants did not adequately encode many of the items in the learning phase.

Participants were marked as correct or incorrect overall for each item they were prompted to recall. To be marked correct overall for a given item, participants must have answered at least two out of three trials for each item correctly. Two was chosen as a cut-off point to minimize data loss while also only including items that participants had repeatedly retrieved. This led to the exclusion of 30.20% of trials in the category-cued-recall condition and 24.44% of trials in the category-stem-cued-recall condition.

2.8.2 Final recall performance.

Category-cued recall. A one-way repeated-measures ANOVA indicated that the proportion of items correctly recalled significantly differed according to the retrieval practice status of the item ($F(1.53, 58.14) = 121.03$, $p < .001$; Huynh-Feldt-corrected degrees of freedom are reported due to violation of the assumption of sphericity: $\chi^2(2) = 15.81$, $p < .001$, $\varepsilon = .77$). We can see from Table 2.2 that recall performance was highest for Rp+ items, which indicates a typical rehearsal effect. Bonferroni-adjusted pairwise comparisons confirmed that participants recalled significantly more of these items compared to baseline ($p < .001$, 95% CI[.36, .55], $d = 1.82$).

RIF was found for Rp- items: Participants recalled fewer Rp- items than Nrp items ($p = .04$, 95% CI[.002, .13, $d = 0.29$). Nonetheless, the RIF effect was small – memory was impaired by just 6%.

Category-stem-cued recall. A one-way repeated-measures ANOVA indicated that the number of items correctly recalled significantly differed according to retrieval practice status ($F(2, 78) = 54.71$, $p < .001$). As expected, participants recalled significantly more Rp+ items, which again simply indicates a rehearsal effect ($p < .001$, 95% CI[.20, .34], $d = 2.13$). However, participants did not show RIF: They retrieved a similar proportion of unpracticed items, regardless of whether those items were related (Rp-) or unrelated (Nrp) to practiced items ($p = 1$, 95% CI[-.05, .10]).

Table 2.2

Mean proportion of items recalled for participants completing category-cued and category-stem-cued recall tests in Experiment 1b as a function of Retrieval Practice Status.

	Mean proportion of items correctly recalled (SD)		
	Nrp	Rp-	Rp+
Category-cued	.32 (.21)	.26 (.21)	.78 (.29)
Category-stem-cued	.66 (.15)	.64 (.20)	.94 (.11)

2.9 Discussion

Using a procedure that mimicked the original retrieval practice paradigm, participants demonstrated RIF only when output interference was not controlled. This suggests that no inhibitory mechanisms were engaged when participants selectively practiced retrieving the category-item pairs and that the memory impairment for Rp- items in the category-cued recall test is merely due to interference created by the order in which participants recalled the items at test. We can therefore be reasonably confident that the lack of RIF in Experiment 1a was due to the stimulus set, rather than some interaction between retrieval processes and deception.

RIF was not found when controlling for output interference in the category-stem-cued recall test. This suggests that the category items are not strongly enough associated with their category to compete for recall and therefore did not require inhibiting. As in Experiment 1a, retrieval practice performance was fairly poor, indicating that the category-item pairs were not well encoded. Given that the pairs were unrelated and abstracted from any context, it is likely that they were not deeply encoded. New stimuli were therefore developed to encourage deeper, semantic processing for Experiment 2.

2.10 Experiment 2

In Experiment 2, participants viewed a mock crime, which should be more meaningful to participants, easier to encode, and more relevant to a deception context. The new mock crime stimuli consisted of pictures of items for memorization together with a label. Images are typically better remembered than words (Maisto & Queen, 1992; McBride & Dosher, 2002). Images are thought to have more distinctive features than words, which encourage deeper, semantic processing (Curran & Doyle, 2011; McBride & Dosher, 2002). Pictorial stimuli with verbal labels were therefore developed to create dual modes of encoding for a stronger memory trace.

Previous research has successfully elicited RIF with pictorial stimuli in the context of eyewitness memory, showing that participants experience RIF when questioned about a portion of an event they previously experienced (MacLeod, 2002; Shaw, Bjork, & Handal, 1995). In one study, for instance, participants were shown images of items allegedly stolen from two houses and were then questioned on half of the items from one of the houses. When participants were later tested on all of the items from both houses, they showed an 11% memory impairment for items that came from the same house as the items they were questioned on, compared to when those items came from the other house (MacLeod, 2002).

The present experiment used a similar procedure to MacLeod (2002), except the framing was reversed: Participants were asked to imagine that they were a perpetrator, rather than a witness, and a fabrication condition was added. The stimulus categories were a crime-relevant image sequence (containing items from the mock crime) and a crime-irrelevant image sequence (containing items from a non-criminal activity), analogous to the two houses used as stimulus categories in MacLeod. Each image sequence contained 10 items for participants to memorize. Participants were then repeatedly questioned on half of the items from one of the image sequences and responded either truthfully (by providing the correct answer) or deceptively (by fabricating their own false answer). After a short delay, participants completed a surprise final recall test for all items initially learnt. It was expected that participants would show better memory for the pictorial mock crime stimuli and that this would bring about the conditions necessary to elicit RIF in both liars and truth tellers.

2.11 Method

2.11.1 Participants.

The experiment was completed in the laboratory by 182 participants in exchange for course credit (mean age = 19.6 years, SD = 2.73 years, range = 18-44; 39 male, 143 female). Most participants were native English speakers (n = 123). Of those remaining, 39 rated their English as excellent, 17 as good, 2 as satisfactory and 1 did not answer. The study was approved by the Psychology Department ethics committee at the University of Warwick.

2.11.2 Stimuli.

The stimuli consisted of 20 still images extracted from two videos. First, the two video sequences were filmed, one for each stimulus category: Crime-relevant and crime-irrelevant. The crime-relevant sequence depicted a gloved person breaking into a laboratory to hack a computer and steal numerous documents, and the crime-irrelevant sequence showed a gloved person performing numerous cleaning duties, such as emptying a bin and vacuuming. The videos were filmed using a GoPro Hero 4 Session with a resolution of 1920x1440. Filming took place in a campus building that students cannot access to ensure that participants would not recognize any of the items or scenes taken from the videos. Each video was filmed from a point-of-view (POV) perspective and depicted a sequence of actions, as well as items associated with those actions (for example, breaking open a door with a crowbar).

Ten still images were then extracted from each video, each depicting one action and its associated item. Participants were instructed to memorize each item. The actions and associated items are described in Appendix 3. Image sequences were used instead of the videos themselves to control the encoding time for each item. Each still image included a caption displayed at the bottom of the screen describing what was happening in the image with the item for memorization written in red. Items for memorization were also presented in the top right corner of the screen to show them clearly and ensure that participants focused on remembering the item specifically. Figure 2.2 shows example images from the crime-relevant sequence.

The first and last items in each image sequence were non-critical items to control for primacy and recency effects and the middle 8 items were critical items for which memory was tested. Each sequence also contained several filler images that did not show any item to be remembered, but were inserted for storytelling purposes. Piloting ensured that baseline memory did not significantly differ across the two sequences (see Appendix 4). Participants completed the experiment in the laboratory on 22-inch monitors with a resolution of 1920x1080.

Figure 2.2. Three images from the crime-relevant image sequence. Participants were instructed to memorize items written in red.

2.11.3 Design and procedure.

The experimental design was the same as Experiment 1a, except the Veracity variable was manipulated between-subjects, such that participants either lied or told the truth in response to all questions in the retrieval practice phase. This allowed a fully factorial

3x2 design and precluded the possibility that task switching between lying and truth telling would obscure any effects of repeating lies on memory. Participants were informed that the study concerned the relationship between intelligence and the ability to construct plausible lies. This was to motivate participants to effortfully construct plausible lies. After providing consent and demographic information, participants began the learning phase.

Learning (perpetration). Participants were asked to imagine that they worked as a cleaner for a company that develops high-tech devices, and that they steal and sell on the company's developments to supplement their income. Participants were informed that they must steal the blueprints for a new device while at work and that they therefore have two tasks: To execute their daily cleaning duties and to break into the office to steal the blueprints. Participants then watched the crime-relevant (theft) and crime-irrelevant (cleaning) image sequences and were instructed to memorize items written in red. Images containing items for memorization were displayed for 8 seconds each and filler images were displayed for 5 seconds each. The total time to watch the image sequences was 4 minutes. All images were presented full screen. The order of the crime-relevant and crime-irrelevant image sequences was counterbalanced across participants.

Retrieval practice (interrogation). Participants were questioned on half of the items from one of the image sequences (4 critical items; for example, "*What was the password to access the IT system?"*). Questions were presented one-by-one in the center of the screen and participants typed their response. Half of the participants were instructed to answer the questions truthfully by reporting the relevant item from the image sequence memorized in the learning phase. The remaining participants were instructed to respond deceptively by fabricating a false answer. Participants responding deceptively were told that they should construct plausible lies that a law enforcement officer would believe. They received 1 minute to answer each question, after which point, the experiment automatically progressed onto the next question.

Each question was asked three times and the question order was predetermined so that retrieval practice occurred on an expanding schedule (as in Anderson et al., 1994). There was an average of 3 trials between the first and second test and an average of 5.25 trials between the second and third test. Within the expanding schedule, the trial order was organized so that there were no recurring patterns of questions. Liars were asked to repeat the same lie each time the relevant question was asked. To ensure that participants understood the task, the first two questions were practice questions that asked about the first and last items in the relevant image sequence, as these were filler items inserted only for controlling for primacy and recency effects. Liars were given example answers and a comprehension check to ensure that they understood how they were expected to answer. For instance, the example of a plausible lie given in response to the question "What was used to get into the office?" was "a number code" (instead of the truthful answer: "a keycard"). The retrieval practice phase therefore consisted of 14 questions in total: 2 practice and 12 critical (1 question for each of the 4 critical items, repeated 3 times).

The items addressed in this phase varied according to 4 counterbalancing sets across participants to ensure that every item served in each retrieval practice condition (Rp+, Rp-, Nrp). Each set contained half of the items from one of the categories (crime-relevant or crime-irrelevant). Items were randomly allocated to each set.

Distractor. This was identical to Experiment 1a, except the tasks were described as intelligence tests to fit with the cover story.

Surprise final recall. Participants were questioned about all 16 critical items from the learning phase and were explicitly instructed to answer all questions truthfully. Questions were shown one-by-one in the center of the screen in a random order across participants to control for output interference. There was no time limit for this task. Items asked about in the retrieval practice phase were cued using the same questions; the phrasing was unchanged. After answering each question, participants were asked to rate their confidence that the item they recalled was the item learnt in phase one on a scale of 0-100, where 0 = *completely uncertain* and 100 = *completely certain*. On completion, participants were thanked and debriefed.

2.12 Results

2.12.1 Retrieval practice performance.

Participants performed well in the retrieval practice phase, indicating that they encoded the mock crime items well and adhered to instructions. Truth telling participants correctly recalled the item in 87.73% of trials (*SD* = 17.76%). Lying participants followed instructions by providing an alternative response in 92.67% of trials (*SD* = 16.79%). The mock crime stimuli therefore led to considerably better performance in the retrieval practice phase than the category-item pairs used in the previous experiments.

Participants were asked about each item three times in the retrieval practice phase. Items were included in the recall analyses only when participants answered all three trials for that item in the retrieval practice phase correctly (7.97% of the recall data for liars and 5.77% of the recall data for truth tellers was excluded based on this criterion).

2.21.2 Final recall performance.

All 182 participants were included in the analysis. A 2x3 mixed ANOVA performed on the proportion of items correctly recalled revealed a significant main effect of Veracity ($F(1, 180) = 10.89$, $p = .001$) and Retrieval Practice Status ($F(1.91, 343.54$[3]$) = 10.62$, $p < .001$). These main effects were superseded by a significant interaction ($F(1.91, 343.54$[3]$) = 6.50$, $p = .002$), thus the effect of retrieval practice on the proportion of items correctly recalled depended on whether participants lied or told the truth. Simple main effects analyses were conducted to follow up this interaction and the results are reported below for liars and truth tellers individually. Table 2.3 shows the descriptive statistics for participants' final recall performance.

Recall performance in liars. As in Experiment 1a, the hypothesis was that repeating lies would promote inhibition of the corresponding truthful item and that this would lead to poorer recall performance for items participants practiced lying about (Rp+ items) compared to baseline (Nrp items). Recall performance was indeed significantly impaired

[3] Huynh-Feldt-corrected degrees of freedom are reported due to violation of the assumption of sphericity: ($\chi^2(2) = 11.93$, $p = .003$, $\varepsilon = .95$).

in liars (*F*(1.87, 167.45) = 5.0, *p* = .009; Huynh-Feldt-corrected degrees of freedom are reported due to violation of the assumption of sphericity: ($\chi^2(2) = 8.99$, *p* = .01, ε = .93). However, the impairment to recall performance was not as predicted: Recall was poorer for Rp- items, rather than Rp+ items, as indicated by the following post-hoc tests.

Repeating fabrications did not affect later memory for the corresponding truthful information. Although Table 2.3 shows that recall performance for Rp+ items is marginally lower than Nrp items, Bonferroni-adjusted pairwise comparisons indicated that this difference was not significant (*p* = 1, 95% CI[-.04, .09].

Nonetheless, practicing lies did impair memory for Rp- items; that is, items that participants did not lie about in the retrieval practice phase, but are related to items they lied about by belonging to the same image sequence. Table 3 shows that participants recalled 10% fewer of these items as a result of practicing their lies. Bonferroni-adjusted pairwise comparisons indicated that this is a significant impairment (*p* = .007, 95% CI[.02, .17], *d* = 0.44). Poorer performance for Rp- items compared to Nrp items mimics the typical RIF effect. Practicing lies therefore did lead to RIF of the truth, but not in the way hypothesized.

Bayesian analysis. As in Experiment 1a, a Bayesian analysis was conducted to determine the likelihood that the null finding reflects no genuine effect, as opposed to a Type II error. A Bayesian one-way ANOVA on the proportion of items correctly recalled for Nrp, Rp- and Rp+ items in liars revealed an overall BF_{01} = 0.28, which provides substantial evidence in favor of the alternative hypothesis that there is a difference in recall performance across the three retrieval practice groups (Jarosz & Wiley, 2014).

Post hoc tests with a corrected prior to adjust for multiple comparisons (as performed by JASP software; [Jasp Team, 2019]) were consistent with the findings of the Frequentist ANOVA, revealing a BF_{01} = 5.72 for the difference in recall performance between Nrp and Rp+ items. This provides substantial evidence for the null hypothesis that there is no difference in recall performance between baseline items and items participants repeatedly lied about. Comparing the difference in recall performance between Nrp and Rp- items revealed a BF_{01} = 0.09, which is considered strong evidence in favor of the alternative hypothesis that repeating lies did impair memory for items related to those lied about (Jarosz & Wiley, 2014). Thus, the Frequentist and Bayesian ANOVA converge on the conclusion that repeating fabrications did not impair memory for the corresponding truthful items, but did impair memory for related items.

Recall performance in truth tellers. The recall performance of truth tellers follows the same pattern of results as in Experiment 1a: Performance was affected by the retrieval practice status of the item ($F(2, 180) = 13.94$, $p < .001$) due to improved memory performance for items truthfully practiced (Rp+ items; $p < .001$, 95% CI [.06, .17], d = 0.58), but not impaired memory for related items that were unpracticed (Rp- items; p = 1, 95% CI [-.07, .06]). Thus, the typical RIF effect was again not replicated. A Bayesian t-test comparing recall performance for Rp- and Nrp items revealed that the data obtained were more than 8 times more likely under the null hypothesis compared to the alternative hypothesis (BF_{01} = 8.62), providing substantial evidence that there is no RIF effect using this procedure (Jarosz & Wiley, 2014).

Table 2.3

Mean proportion of items correctly recalled and mean confidence ratings (scale 0-100) in the final recall test in Experiment 1a as a function of Retrieval Practice Status and Veracity.

	Proportion recalled (SD)			Confidence (SD)		
	Nrp	Rp-	Rp+	Nrp	Rp-	Rp+
Liars	.79 (.18)	.69 (.27)	.76 (.25)	79.38 (14.65)	76.65 (21.53)	79.66 (18.78)
Truth Tellers	.79 (.21)	.79 (.24)	.90 (.17)	81.21 (14.41)	82.81 (15.66)	86.35 (15.71)

Note: "Rp+" represents practiced items (practiced lies or practiced truths), "Rp-" represents unpracticed items that are related to practiced items, and "Nrp" represents unpracticed items that are unrelated to practiced items and therefore serves as the baseline to which other groups are compared.

2.12.3 Confidence.

As in Experiment 1a, Participants' confidence in their memories was assessed to determine if practicing lies led to more subtle changes in memory than impairing explicit recall. Confidence ratings for trials where participants did not provide an answer or answered "don't know" were excluded from the confidence analysis, which led to the removal of 10.77% of the data (truth tellers = 5.0%, liars = 5.77%).

Table 2.3 shows participants' mean confidence ratings (on a scale of 0 = *completely uncertain* to 100 = *completely certain*). Confidence was generally high and the differences between conditions were very small. A 2x3 mixed ANOVA revealed a main effect of Veracity: Truth tellers were overall more confident than liars (M_{Truth} = 83.40, *SD* = 15.38; M_{Liars} = 78.56, *SD* = 18.52; $F(1,180) = 5.34$, $p = .02$; 95% CI[0.70, 9.0], d = 0.28). Confidence was also affected by the retrieval practice status of the item, indicated by a main effect of Retrieval Practice Status ($F(2, 360) = 4.25$, $p = .02$). Specifically, participants were more confident in their memories for Rp+ items compared to Nrp items ($M_{\text{Rp+}}$ = 83.0, *SD* = 17.59; M_{Nrp} = 80.21, *SD* = 14.52; $p = .05$, 95% CI[0.03, 5.56], $d = 0.17$), but there was no difference in confidence ratings between Rp- items and Nrp items ($p = 1$, 95% CI[-2.55, 3.51]), as indicated by Bonferroni-adjusted pairwise comparisons. Although liars' showed impaired recall for items related to those that they lied about (Rp- items), there was no corresponding decrease in their confidence: The interaction between Veracity and Retrieval Practice Status was not significant ($F(2, 360) = 2.60$, $p = .08$). Thus, there was no evidence of RIF in participants' confidence ratings.

2.13 Discussion

As in Experiment 1a, repeating fabrications did not impair recall of the corresponding truthful item. However, there was RIF of the truth, but not in the way predicted: Memory was impaired for items that were not prompted in the retrieval practice phase but belonged to the same category as lied-about items.

RIF of items related to those lied about is a surprising finding, especially since RIF was not found in truth telling participants. The effect is therefore specific to fabrication. One explanation for this finding is that the act of generating a plausible alternative answer initiated a retrieval search that activated other items from the same image sequence. Items from the same image sequence might achieve threshold activation quickly because they were recently studied. Indeed, one prominent theory of the cognition of deception—Activation Decision Construction Action Theory (ADCAT; Walczyk et al., 2014)—states that recently encoded memories are the first to be activated in the construction of plausible lies. Since items from the image sequence could not be used as a fabrication, their activation may have interfered with the retrieval of useful alternatives, which in turn may have initiated the inhibitory processes that promote forgetting. Nonetheless, this raises the question of why the truthful item was not also activated and in turn inhibited. It is therefore unclear what underlies this effect.

One reason why RIF was not found for the corresponding truthful items might be that the association between the studied items and their cues was still too weak, despite efforts to promote better encoding. In standard RIF studies, there are two considerations regarding associative strength: [1] the strength of the association between the cue (e.g., "Fruit-Or__") and the recall target (e.g. "orange") and [2] the strength of the association between the cue and the associated non-targets/competitors (e.g. "lemon"). RIF only occurs if the association between the cue and the competitors is equal to or stronger than the association between the cue and the target, since it is only under these conditions that there is competition for recall and therefore a retrieval search for the target (Anderson, 2003; Anderson et al., 1994). If the association between the cue and target outweighs that between the cue and the competitors, the target is quickly and easily retrieved without any interference from associated items, obviating the need to inhibit them (Norman, Newman, & Detre, 2007).

For the present experiments, the retrieval target is the lie and the associated non-target is the truth. Thus, for the truth to be a competitor, the association between the cue and the truth must be equal to or stronger than the association between the cue and the lie. If not, the lie may be retrieved without any interference from the truth. Because participants lied about recently learnt material, it is likely that their memory for the truthful answer was simply not strong enough to compete for recall with the lie. Participants encountered each item only once in the learning phase but delivered their lie in the subsequent retrieval practice phase three times. This may have led to a stronger, and more recent, memory of the lie than the truth such that the lie was quickly and easily retrieved without a competing truthful response. A different approach is therefore required so that participants possess strong memories of the truth to create the conditions necessary to see RIF. This interpretation is supported by the further finding that the standard RIF effect was again not replicated in truth tellers.

According to ADCAT, we require inhibition of the truth to a greater extent when the truth is well-rehearsed, or "entrenched" because the truth is more likely to be the automatic dominant response in such cases (Walczyk et al., 2014). In fact, ADCAT hypothesizes that "entrenched truths (e.g., those central to respondents' lives, deeply held beliefs) will cause more proactive interference with lying, that must be inhibited, than will peripheral or recently encoded truths" (Walczyk et al., 2014, p. 32). It is therefore likely that repeating a lie might only lead to RIF of the truth when the truth is entrenched, rather than recently learnt (as in Experiments 1a and 2). Consistent with this reasoning, one study showed that liars were more effectively discriminated from truth tellers when they lied about details pertaining to their everyday lives, and are therefore entrenched in memory, compared to when they lied about recent autobiographical events (Walczyk et al., 2009). This suggests that lying about entrenched truths does indeed generate more proactive interference when lying, which in the present context might create the retrieval competition necessary to see RIF of the truth. Experiment 3 therefore manipulates the strength of the truth in memory in a final attempt to determine if there are inhibitory processes engaged during fabrication that can promote RIF of the truth.

2.14 Experiment 3

An obvious way to manipulate the accessibility of the truth in memory is to include a training session whereby participants strengthen their memory for a subset of items. However, a training session is essentially a form of retrieval practice, which could obscure any further effects of selective practice in the retrieval practice phase. This experiment could not therefore use a training session to increase accessibility or any stimuli that participants would be encountering for the first time (such as a mock crime). Instead, stimuli that capitalize on participants' existing knowledge was used to ensure that high accessibility items were indeed entrenched in memory. Experiment 3 therefore returned to verbal stimuli to allow greater control of each item's accessibility.

Similar to Experiment 1a, the stimuli were composed of items that belong to a category. However, instead of learning category-item pairs, participants learnt question-answer pairings (e.g., instead of "Country-Japan", participants learnt "Q: What country does Sushi come from? A: Japan"). Each item therefore formed an answer to a question. This meant that all participants could learn the same items, while allowing their accessibility to be manipulated across participants by changing the question difficulty with which the item is associated. Thus, for one participant, the item "Japan" should be highly accessible when cued with the easy question "What country does sushi come from?", but for another participant, "Japan" should be less accessible (if at all) when cued with the hard question "What country consists of over 6800 islands?". This design therefore eliminated item effects to ensure that any difference in memory between the low and high accessibility conditions is indeed attributable to the item's strength in memory and not because one group of items is inherently better remembered than the other.

Because entrenched truths are unlikely to be forgotten, a reaction-time (RT)-based recognition test was used that could detect more subtle changes in an item's accessibility after lying, rather than a recall test. Past research has shown that RIF is apparent in RT-based recognition tests, as well as recall tests. This manifests as slower RTs to Rp- items than Nrp items, indicating reduced accessibility for Rp- items (Perfect, Moulin, Conway, & Perry, 2002; Veling & van Knippenberg, 2004). Furthermore, the use of RTs in detecting deception is increasingly popular and these tests consistently show that liars are slower to respond to crime-relevant items than truth-tellers (Verschuere et al., 2018). If repeating lies reduces the accessibility of truthful items in memory, participants undergoing these tests might experience less interference from the truth and respond faster to crime-relevant items after practicing their fabrications. This is unlikely to change the outcome of the test, but it may create noise and reduce the test's diagnosticity. Prior research has shown that RTs can indeed change from repeated lying (Van Bockstaele et al., 2012). Thus, Experiment 3 investigates whether repeating fabrications can impair item recognition, as indicated by slower RTs.

The hypothesis was that RIF would occur for items that are highly accessible in memory (i.e., items cued by easy questions), as these items should be strongly associated with their retrieval cue. For truth tellers, the standard RIF effect was predicted (longer RTs to Rp-Truth items). For liars, highly accessible truths should 'pop' into people's minds on receiving the question, interfere with lie generation, and therefore require inhibiting. It was therefore predicted that liars would show longer RTs to lied-about items (Rp+Lie) in the recognition test. Longer RTs for items related to those lied about (Rp-Lie) were also predicted to replicate the RIF effect found in Experiment 2.

2.15 Method

2.15.1 Participants.

The experiment was completed in the laboratory by 211 participants (mean age = 21.14 years, *SD* = 5.02 years, range = 18-50; 139 females, 72 males) in exchange for course credit or a small monetary reward. Most participants were native English speakers (n = 153). Of those participants whose first language was not English, 44 participants rated their English as excellent and the remaining 14 participants rated their English as good or satisfactory. The study was approved by the Psychology Department ethics committee at the University of Warwick.

2.15.2 Stimuli and apparatus.

The stimuli were taken from the same published word norms as Experiment 1a (Van Overschelde et al., 2004). Forty-eight items from 6 different categories (8 per category) were selected as the critical items. These items formed the answers in the question-answer pairings. An additional 4 items from 2 further categories were selected to use as fillers, yielding a total of 56 items.

The categories were semantically and phonetically unrelated to one another to prevent cross-category interference. The items were selected to have high taxonomic strength to ensure that they were clearly associated with their category. The 7 highest-ranking items from each category were selected with the constraint that no two items from a single category could start with the same letter. The items had a mean ranking of 4.63 (*SD* = 0.62; Overschelde et al., 2004).

Each item formed the answer to 2 questions—one easy and one hard. This was to manipulate the accessibility of the items in memory. Easy questions were designed to cue the answer quickly (high accessibility), whereas hard questions were designed so that participants either would not know the answer or would have to think harder to retrieve it (low accessibility). An online pilot questionnaire was conducted to ensure that this manipulation was successful, in which participants (n=56) were asked the easy or hard questions for each item and typed their answers. Participants' recall performance and RT to respond was used to assess the question difficulty. Participants recalled all items more often in response to easy compared to hard questions. The proportion of participants who provided the correct answers in the easy condition (*M* = 0.92, *SD* = 0.11) was significantly greater than the proportion of participants who provided correct answers in the hard condition (*M* = 0.26, *SD* = 0.03; t(55) = 46.57, $p < .001$, 95% CI[0.63, 0.67]). Furthermore, when participants answered correctly for hard questions, they did so significantly slower than for easy questions (M_{hard} = 15.52s, *SD* = 20.88s; M_{easy} = 8.32s, *SD* = 3.03s; t(55) = -2.66, $p = .01$, 95% CI[-12.63s, -1.77s], *d* = 0.48). We can therefore be confident that items were significantly more accessible when cued with easy questions compared to when cued with hard questions. The stimuli are listed in Appendix 5. Participants completed the experiment on 22-inch monitors with a resolution of 1920x1080. The experiment was programmed using PsychoPy (Peirce, 2007).

2.15.3 Design and procedure.

The experiment had a 2 (Veracity: lie, truth) x 2 (Accessibility: low, high) x 3 (Retrieval practice status: Rp+, Rp-, Nrp) mixed design. Veracity was the between-subjects variable. The experiment consisted of 4 phases: Learning, retrieval practice, distractor and a surprise final recognition test. The dependent measure was the RT to identify an item as old or new (learnt or not learnt in the learning phase respectively) in the final recognition test. Figure 2.3 shows an overview of the experimental procedure.

Learning. Participants were told that the experiment concerned the role of memory and reasoning skills in lie construction. After providing consent and demographic information, participants began the learning phase, in which the 56 (48 critical, 8 filler) question-answer pairings were presented, together with the category to which they belong, one-by-one in the center of the screen (see Figure 2.3 for examples). Participants were instructed to learn the answer to each question and were told that their memory would be tested.

Each question began with the category name (e.g. 'What fruit…', 'What country…') so that all questions unambiguously belonged to their category and to prime participants with the category. For each question-answer pairing, the category name and question first appeared for 2.5 seconds, followed by the answer, which remained on screen for 5s. The presentation of the answer was delayed in this way to ensure that participants read the questions and did not only focus on the answers. There was a 0.5s interval between each pairing.

The question answer pairings were presented in a blocked semi-random order. The first and last 4 pairings were filler items to control for primacy and recency effects. For critical pairings, there were 8 blocks each containing one question-answer pairing per category (and therefore 3 easy and 3 hard pairings). The order of the pairings within each block and the block order were randomized across participants. There was a 5 second break in-between blocks.

As previously stated, all participants learnt the same 56 items, but the question difficulty was varied to manipulate accessibility so that category and item effects were eliminated in the final recognition test. Participants always learnt 3 easy (high accessibility) categories and 3 hard (low accessibility) categories. To ensure that all items served in the low and high accessibility conditions, two counterbalancing sets were created so that half of the participants received easy question-answer pairings for 3 of the 6 categories and the remaining half received hard question-answer pairings for those same categories.

Retrieval practice. Participants answered half of the questions for 2 of the 3 easy categories and half of the questions from 2 of the 3 hard categories. They were therefore questioned about 16 items in total (8 easy and 8 hard). Questions were presented one-by-one in the center of the screen.

Half of the participants answered the questions truthfully and the remaining half answered deceptively. Deceptive participants were asked to construct plausible fabrications in response to the questions so that somebody who didn't know the truthful answer would believe their lie. Truthful participants responded with the answer they learnt in the learning phase. Participants had 12 seconds to write their answer. If they did not respond within 12 seconds, the experiment moved onto the next question. Each question was asked 3 times, yielding 48 questions in total. The order of questions was predetermined in the same way as Experiment 2.

The questions varied according to 6 counterbalancing sets. Each set contained half of the items from 2 of the 3 easy categories and 2 of the 3 hard categories. Thus, every item served in all retrieval practice conditions (Rp+, Rp-, Nrp).

Distractor. This was identical to Experiment 2.

Final recognition test. The recognition test was based on the RT-based test used by Veling and van Knippenberg (2004). The test consisted of 10 practice and 96 critical trials, in which the 48 critical items from the learning phase and 48 new items were presented one-by-one in the center of the screen. New items were selected from a further 6 categories from Overschelde et al. (2004) using the same criteria that were used to select the critical items.

Each trial began with an asterisk in the center of the screen, followed by a word after a random interval between 1.5 and 2 seconds. Participants were instructed to indicate whether they learnt the word in the learning phase and to press the “A” key to indicate “yes” and the “L” key to indicate “no” as quickly and accurately as possible. The word remained on screen until keypress. The response options were counterbalanced across participants.

The first 10 trials were practice trials consisting of the 5 of the filler items from the learning phase and 5 new items. Participants were provided with feedback after each trial to familiarize them with the task. On completion of the practice trials, participants were informed that they would no longer receive feedback on their performance. Feedback was not provided for critical items, so that participants could not use this to aid their memory in subsequent trials. On completion of the recognition test, participants were thanked and debriefed.

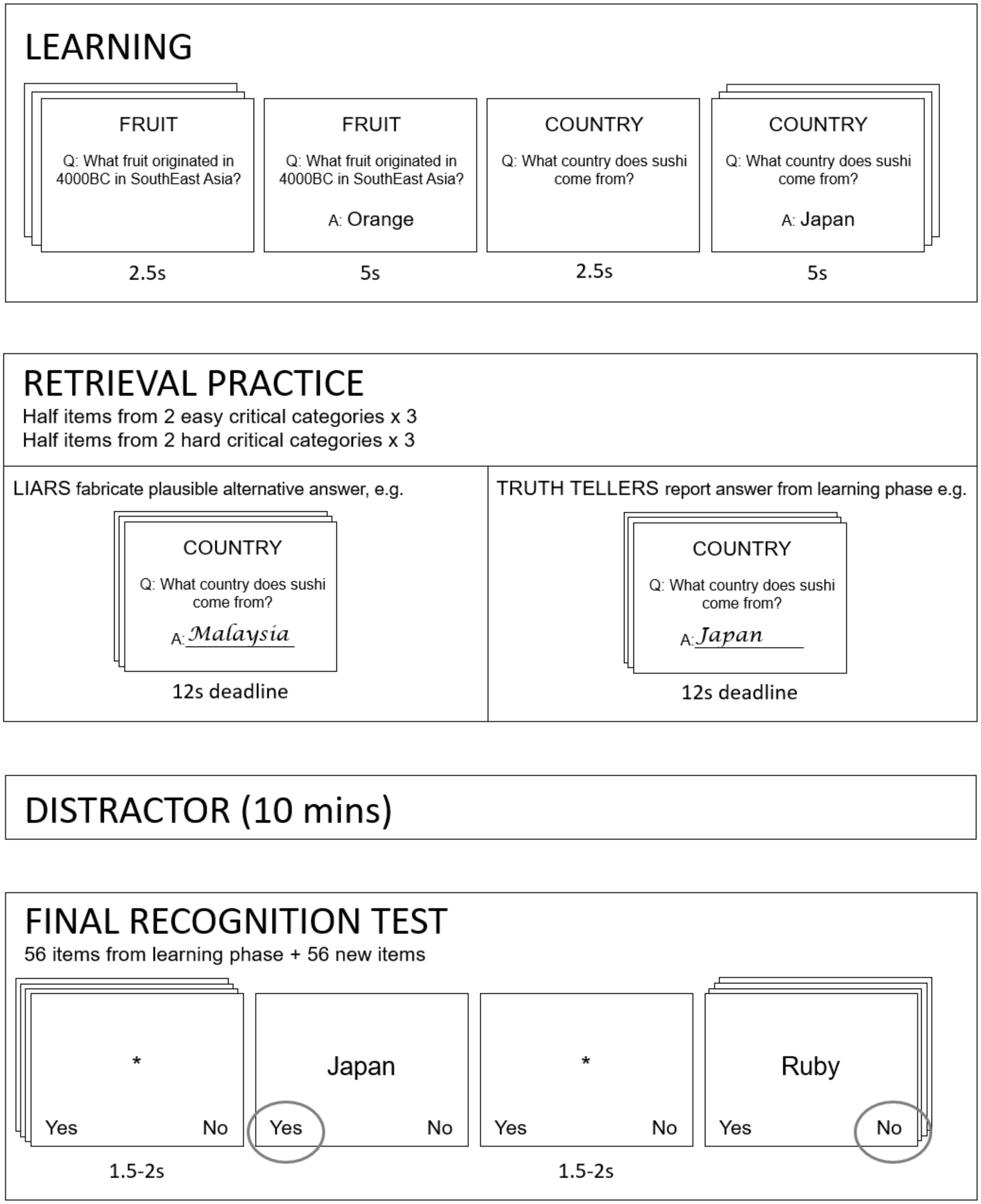


Figure 2.3. Procedural overview of Experiment 3.

2.16 Results

2.16.1 Retrieval practice performance.

Ten participants failed to answer any of the questions correctly and were excluded from all subsequent analyses. Analyses are therefore based on data from 201 participants.

Participants performed well in the retrieval practice phase, indicating that they sufficiently encoded the items in the learning phase. Truth telling participants correctly recalled the item in an average of 83.97% of trials (*SD* = 36.69%). Lying participants correctly provided an alternative response in an average of 82.88% of trials (*SD* = 37.67%).

Participants were asked about each item three times in the retrieval practice phase. To be scored as correct overall for each item, they had to answer at least two trials for each item correctly. Two correct trials was chosen as a cut-off point to minimize data loss while also only including items for which participants had repeated their lies. The corresponding recognition data for items scored as incorrect overall were excluded from all subsequent analyses. This led to the removal of 12.75% of the recognition data.

2.16.2 Recognition performance.

RTs above and below 2 standard deviations of each participant's mean RT per condition were excluded. This led to the removal of less than 0.1% of the data. A multilevel model was fitted to RTs for correct trials only to assess recognition performance.

Multilevel linear models. A multilevel linear model can be likened to a regression model that accommodates within-subject variables by explicitly modelling clustered variables (Field, Miles, & Field, 2012). Thus, a multilevel model is a regression model that can be used for within-subjects designs. Specifically, a multilevel model accounts for the fact that RTs for each level of the Item Accessibility and Retrieval Practice Status variables come from the same participants and are therefore correlated. A multilevel model was used instead of a repeated-measures ANOVA, as the computer randomization led to an unbalanced design of 124 liars and 77 truth tellers. Multilevel models are preferable to repeated-measures ANOVA when data are unbalanced or missing because they do not require whole cases to be removed or data to be imputed where there are missing data, but instead can model parameters based on the available data (Field et al., 2012; Maas & Hox, 2005). Multilevel models are therefore used throughout this thesis instead of repeated-measures ANOVA wherever there is an unbalanced design or missing data. All multilevel models throughout this thesis were conducted in R (R Core Team, 2018) using the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2020).

Model fitting. First, the fitting of the model to the data is described and the results are then interpreted. Figure 2.4 shows that RTs differed depending on the retrieval practice status of the item—including Retrieval Practice Status significantly improved the model fit compared to an intercept-only model ($\chi^2(7) = 33.67$, $p < .001$). Adding the main effect of Item Accessibility did not further improve the model fit ($\chi^2(8) = 0.10$, $p = .75$), but including the main effect of Veracity did ($\chi^2(9) = 6.20$, $p = .01$).

Adding the interaction term between Retrieval Practice Status and Item Accessibility did not improve the model fit ($\chi^2(11) = 5.72$, $p = .06$) at the alpha level of .05, but the model was significantly improved when including the interaction term between Retrieval Practice Status and Veracity ($\chi^2(13) = 25.46$, $p < .001$). The three-way interaction between Retrieval Practice Status, Item Accessibility and Veracity did not improve the model fit ($\chi^2(16) = 0.10$, $p = .95$). The three-way interaction was therefore removed from the model. Because the interaction between Retrieval Practice Status and Item Accessibility just missed the 0.05 significance level, it was kept in the model to further explore the interaction.

Model interpretation. A typical RIF effect is defined by longer RTs to Rp- items than Nrp items, reflecting reduced accessibility of Rp- items. The primary question of this experiment was whether repeating fabrications leads to RIF only when the truth is highly accessible in memory. Support for this hypothesis requires a significant three-way interaction between Retrieval Practice Status, Item Accessibility and Veracity, or at least a significant interaction between Retrieval Practice Status and Item Accessibility. Since the fit of the multilevel model was not improved when including any interaction terms containing Item Accessibility, this hypothesis was not supported: Item Accessibility did not significantly affect RTs.

However, we can see from Figure 2.4 that although Item Accessibility did not significantly predict RTs, the overall pattern of results suggests that it did have some effect on participants' performance. Interestingly, the pattern shown is the opposite to that predicted. It was predicted that liars would take longer to respond to items that they had practiced lying about (Rp+ items) and related items (Rp- items) compared to baseline (Nrp items), but only for high accessibility items. In fact, the direction of the results shown in Figure 2.4 suggests RIF only for low accessibility items.

The model contrasts suggest a potential RIF effect for low accessibility items. Contrasts were set to compare both Rp groups to the baseline Nrp group for low and high accessibility items, collapsed across the Veracity condition. These revealed that the difference between RTs for Nrp and Rp- items was greater for low accessibility (M_{diff} = -39.28, SD = 281.80) compared to high accessibility items (M_{diff} = 11.56, SD = 282.89), but this difference was not significant at the alpha level of .05 (b = 50.84, $t(590) = 1.82$, $p = .07$). The difference in RTs between for Nrp and Rp+ items did not significantly differ for low accessibility (M_{diff} = 100.80, SD = 320.16) and high accessibility items (M_{diff} = 102.29, SD = 384.12; b = -13.62, $t(590) = -0.48$, $p = .63$).

The question is therefore whether the longer RTs for low accessibility Rp- items merely reflects a chance finding or if the non-significance is due to a Type II error from a lack of statistical power. A Bayesian analysis was conducted to determine which of these possibilities is most plausible. A Bayesian one-way ANOVA comparing RTs for Nrp, Rp- and Rp+ items for low accessibility items revealed a $BF_{01} < .0001$, which is considered decisive evidence in favor of the alternative hypothesis that there is a difference between retrieval practice groups for low accessibility items (Jarosz & Wiley, 2014). Post hoc tests with a corrected prior to adjust for multiple comparisons (as performed by JASP software; [Jasp Team, 2019]) revealed a $BF_{01} = 1.26$ for the difference in RTs between Nrp and Rp- low accessibility items, which is considered only anecdotal evidence in favor of the null hypothesis (Jarosz & Wiley, 2014). This suggests that the non-significant difference between RTs for low accessibility Nrp and Rp- items may be a Type II error due to a lack of statistical power. Consistent with the multilevel model contrasts, the comparison of RTs for low accessibility Nrp and Rp+ items revealed a BF_{01} = 0.02, which again provides only anecdotal evidence in favor of the null hypothesis and suggests that RTs may indeed differ for these groups.

As in Experiments 1a and 2, a rehearsal effect was found in truth tellers, represented by the significant interaction between Retrieval Practice Status and Veracity. This interaction is depicted in Figure 2.5, which shows that truth tellers responded more quickly to Rp+ items than liars. The model contrasts revealed that the difference in RTs between Nrp and Rp+ items is significantly greater in truth tellers (M_{diff} = 141.86, *SD* = 159.22) than liars (M_{diff} = 61.14, *SD* = 328.08; *b* = -125.34, $t(393) = -4.32$, $p < .001$). This reflects stronger memory from truthfully rehearsing the items in the retrieval practice phase. The difference in RTs between Nrp and Rp- items did not significantly differ for liars (M_{diff} = -11.88, *SD* = 181.73) and truth tellers (M_{diff} = -17.04, *SD* = 225.74; *b* = 5.16, $t(393) = 0.18$, $p = .86$). In sum, when not controlling for the accessibility of the truth, repeating lies did not affect later memory for items learnt in the learning phase, but telling the truth improved memory, due to a standard rehearsal effect.

2.16.3 Recognition results summary.

Taken together, these results indicate that manipulating the accessibility of the truth in memory did not influence the RIF effect in this experiment. However, the results of the Bayesian analysis suggest that there may indeed be an RIF effect for low accessibility items, but that this effect was missed due to a lack of statistical power. Nonetheless, the present data do not provide sufficient evidence that repeating lies promotes RIF of the truth via an inhibitory mechanism; participants were not significantly slower to respond to items that they had practiced lying about (Rp+ items) compared to baseline items for neither low nor high accessibility items. Additionally, the standard RIF effect in truth tellers was once again not replicated: Participants were not slower to respond to items related to those they had practiced retrieving (Rp- items) compared to baseline items for low or high accessibility items.

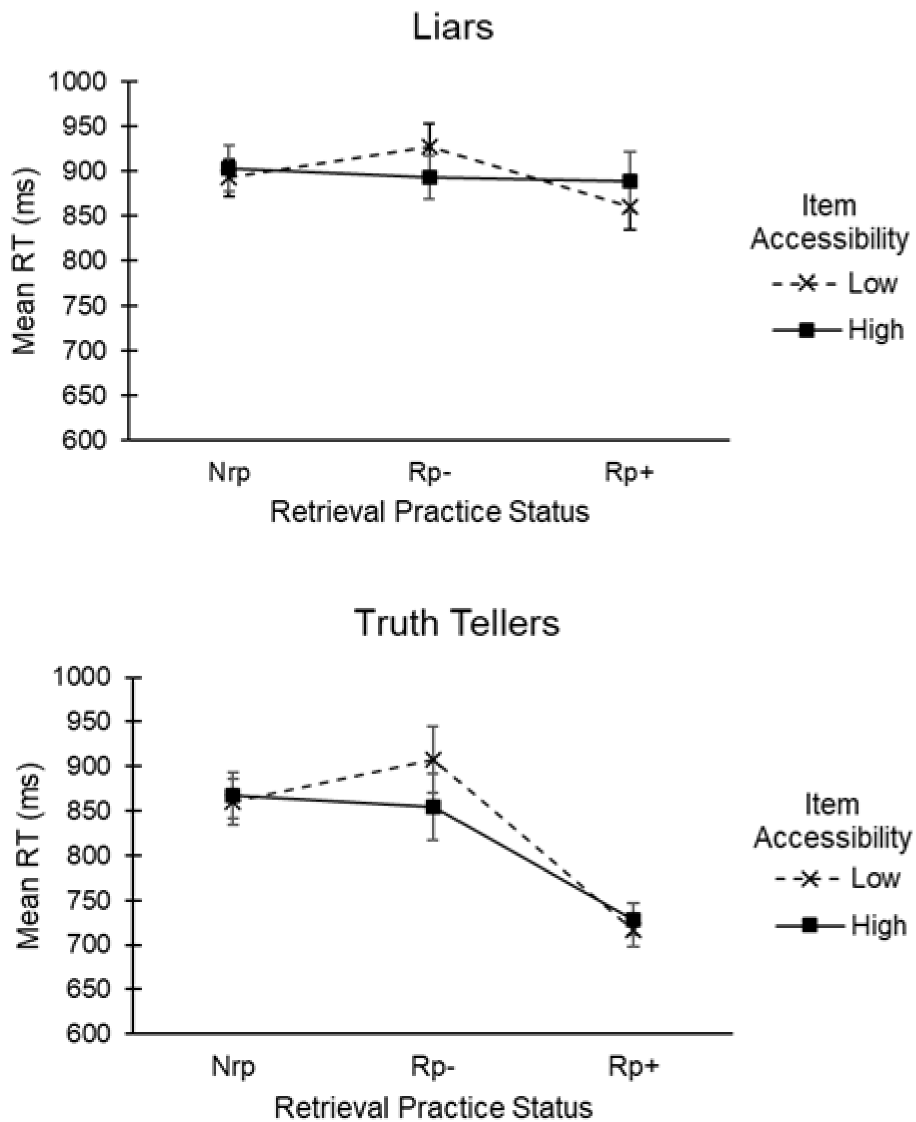


Figure 2.4. Mean reaction time (ms) for liars (top panel) and truth tellers (bottom panel) to indicate whether an item was learnt in the learning phase as a function of Retrieval Practice Status and Item Accessibility. Error bars indicate ±1 SE of the mean.

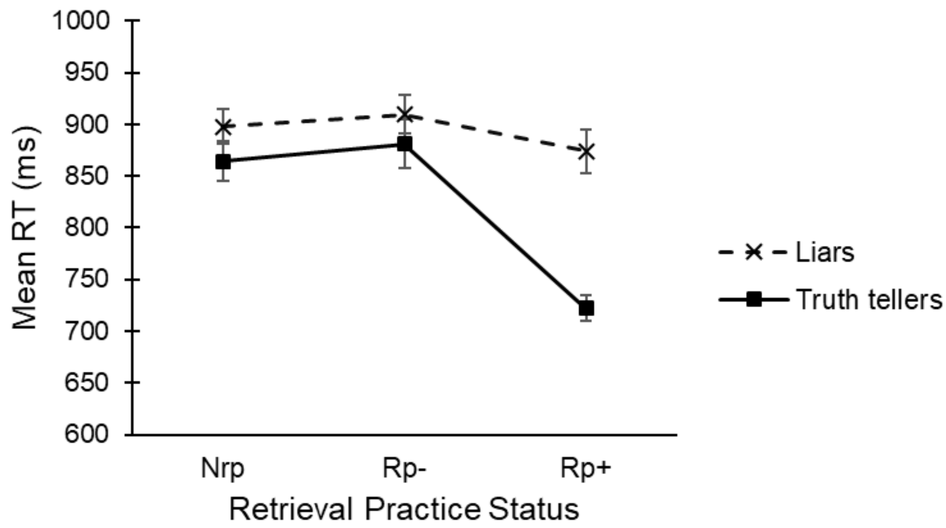


Figure 2.5. Mean reaction time (ms) for liars and truth tellers to indicate whether an item was learnt in the retrieval practice phase as a function of Retrieval Practice Status (collapsed across Item Accessibility). Error bars indicate ±1 SE of the mean.

2.16.4 Error rates.

Participants' error rates supplement the RT analyses, as any difference in errors across conditions compared to baseline may arise from changes in memory attributable to selective retrieval practice. An error constitutes participants incorrectly responding "no" that the item shown was not studied. The error rates therefore provide an estimate of the rate of omission errors for each type of item. A multilevel model was fitted to the error rate data in the final recognition test.

Model fitting. As for the recognition analyses, the fitting of the model is described first and the model results are then interpreted. Error rates differed depending on the retrieval practice status of the item—including Retrieval Practice Status as a predictor significantly improved the model compared to an intercept-only model ($\chi^2(7) = 39.52$, $p < .001$). Adding the main effect of Item Accessibility did not further improve the model fit ($\chi^2(8) = 1.73$, $p = .18$) or Veracity ($\chi^2(2) = 21.5$, $p < .001$), but the main effect of Veracity did ($\chi^2(9) = 16.26$, $p < .001$).

Figure 2.6 shows that there was a clear interaction between all variables. Indeed, the model fit was improved with the addition of the interaction between Retrieval Practice Status and Item Accessibility ($\chi^2(11) = 18.81$, $p < .001$), Retrieval Practice status and Veracity ($\chi^2(13) = 60.66$, $p < .001$), and the three-way interaction between Retrieval Practice Status, Item Accessibility and Veracity ($\chi^2(16) = 22.83$, $p < .001$).

To break down this three-way interaction, separate multi-level models were fitted to truth tellers' and liars' error data. For liars, including the main effect of Retrieval Practice Status significantly improved the model fit compared to an intercept-only model ($\chi^2(7) = 9.21$, $p = .01$), but the main effect of Item Accessibility did not further improve the model fit ($\chi^2(8) = 2.65$, $p = .10$). Including the interaction term between Retrieval Practice Status and Item Accessibility significantly improved the model fit ($\chi^2(10) = 35.16$, $p = < .001$).

For truth tellers, including the main effect of Retrieval Practice Status significantly improved the model fit compared to an intercept-only model ($\chi^2(7) = 103.25$, $p < .001$), but neither the main effect of Item Accessibility ($\chi^2(8) = 0.01$, $p = .93$) nor the interaction between Retrieval Practice Status and Item Accessibility ($\chi^2(10) = 1.71$, $p = .43$) significantly improved the model fit.

Model interpretation. First, for liars, the top panel of Figure 2.6 reveals that the accessibility of the item influenced error rates only for items lied about (Rp+ items). Error rates were higher for low accessibility items that were lied about compared to high accessibility items. The model contrasts confirmed that the difference in error rates between Nrp and Rp- items did not significantly differ between low and high accessibility items ($b = -0.05$, $t(367) = -1.72$, $p = .09$). However, repeatedly lying about an item (Rp+ items) significantly increased errors when those items were less accessible compared to when they were more accessible ($b = 0.11$, $t(367) = 4.16$, $p < .001$).

For truth tellers, the bottom panel of Figure 2.6 suggests that the accessibility of the items did not influence the effect of retrieval practice on error rates (confirmed by the non-significant interaction between Retrieval Practice Status and Item Accessibility).

Unsurprisingly, truth tellers made significantly fewer errors for items they had practiced retrieving in the retrieval practice phase (Rp+ items) compared to Nrp items (b = -0.24, $t(152)$ = -8.65, p < .001). There was no difference in error rates between Rp- items and Nrp items (b = -0.02, $t(152)$ = -0.79, p = .43). In sum, truth tellers made fewer errors if they practiced retrieving the item in the retrieval practice phase.

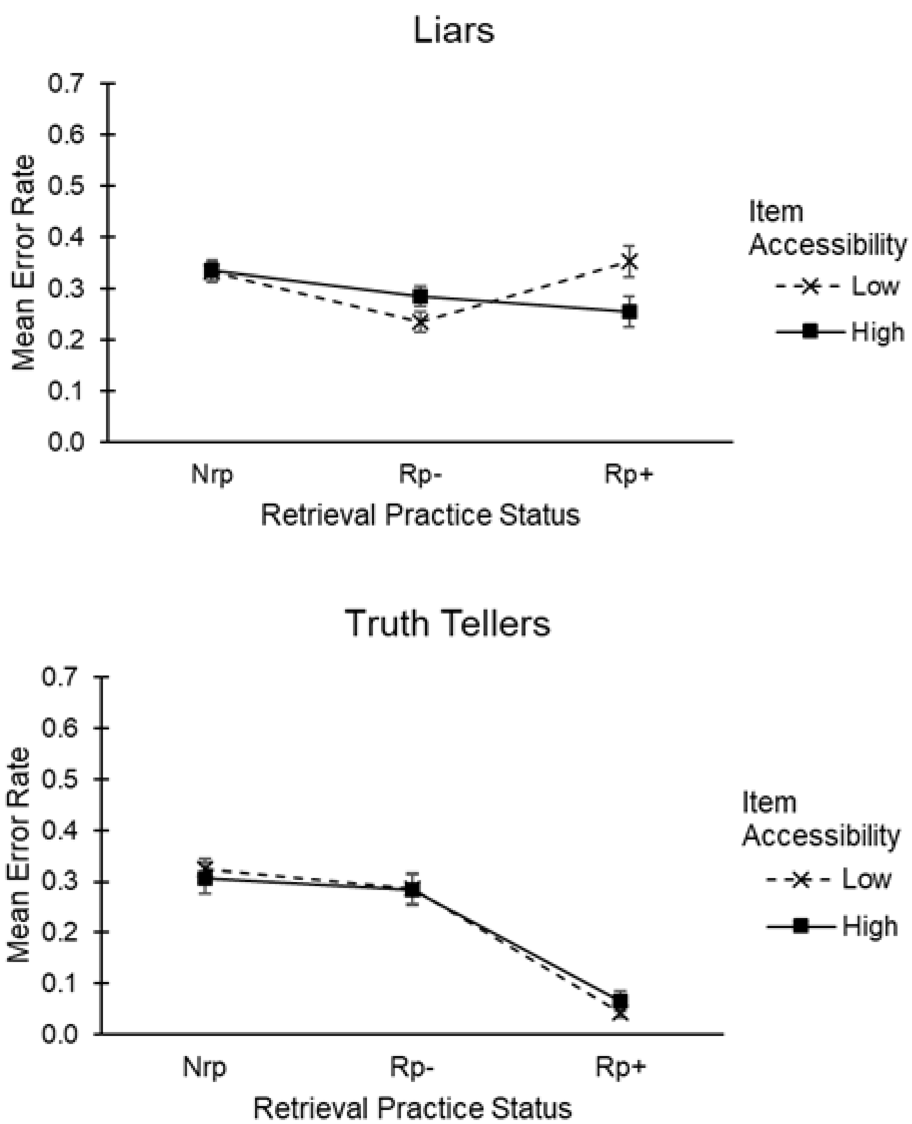


Figure 2.6. Mean proportion of errors in the RT-based-recognition test for liars (top panel) and truth tellers (bottom panel) as a function of Retrieval Practice Status and Item Accessibility. Error bars represent ±1 SE of the mean.

2.16.5 Error rate results summary.

The accessibility of the truth in memory affected liars' error rates, but not truth tellers'. For liars, error rates were increased when participants repeated their lies for items that were less accessible in memory (cued by hard questions) compared to items that were more accessible in memory (cued by easy questions). This suggests that an individual who repeatedly lies about information that is weaker in memory might make more omission errors later on than if they had not lied. Conversely, truth tellers' error rates were lower for items they had practiced truthfully retrieving in the retrieval practice phase, regardless of the item's accessibility in memory (whether it was cued by an easy or hard question).

2.17. Discussion

The goal of Experiment 3 was to determine if repeating fabrications leads to RIF of the corresponding truthful information only when the truth is highly accessible or "entrenched". If so, this would suggest that the null findings of the previous experiments may be due to the accessibility of the truth being too low to create the conditions that lead to RIF. It was predicted that RIF would occur only for highly accessible items and therefore that RTs would be slower for items participants repeatedly lied about (RP+Liars) and for items related to those lied about (Rp-Liars) compared to baseline (Nrp).

The hypothesis was not supported. The accessibility of the truth did not significantly affect RTs. The null findings in the previous experiments were therefore unlikely to be the result of weak memory for the truth. In fact, contrary to the hypothesis, the results suggested that RIF may be more likely to occur for weaker items in memory in both liars and truth tellers (although this finding was not statistically significant, the Bayesian analysis indicated that it may be a missed effect due to a lack of statistical power).

Interestingly, participants made fewer omission errors after repeating fabrications for high accessibility items compared to when they repeated fabrications for low accessibility items. This suggests a slight rehearsal effect: Repeating lies may slightly facilitate memory when the truth is highly accessible. This slight rehearsal effect was not reflected in participants' RTs, but this could be because RTs were only analyzed for correct trials. Thus, when participants successfully remembered the item studied, there were no further differences in the ease of retrieval of high and low accessibility items studied.

One reason that lying about high accessibility items decreased omission errors could be that highly accessible truths were indeed automatically activated and therefore covertly rehearsed. Interestingly, the facilitative effect on memory suggests that the truth was not activated to the extent that it interfered with lie construction and required inhibiting. The implications of this possibility are discussed further in the next section.

2.18 General Discussion

2.18.1 Summary of main findings

The experiments reported in this chapter aimed to determine if repeating fabrications can promote RIF of the truth via an inhibitory mechanism. Given that the cognitive processes that lead to RIF also appear to be recruited when people fabricate, it was predicted that repeating fabrications would promote inhibition of the corresponding truthful item in memory and that this would impair later recall of the truth.

Experiments 1a and 2 investigated the effect of repeating fabrications on recall, whereas Experiment 3 investigated its effects on recognition. Contrary to the hypothesis, repeating fabrications did not impair recall or recognition of the corresponding truthful information and there was no evidence of impaired memory in participants' confidence ratings. Thus, no evidence was found that repeating fabrications can impair memory for the corresponding truthful information via RIF.

Repeating fabrications did, however, promote RIF of related truthful information that was not probed in the retrieval practice phase. In Experiment 2, repeating fabrications led to an absolute memory impairment of 10% and a relative impairment of 13% – a small to medium effect size consistent with the RIF literature (Murayama et al., 2014). A small, but non-significant, impairment for these items was also found in Experiment

1a, however the Bayesian analyses suggested that this non-significant finding could be a missed true effect. This may be due to a lack of statistical power arising from the exclusion of recall data for items participants did not correctly respond to in the retrieval practice phase. Similarly, the overall pattern of results in Experiment 3 suggested a small, but non-significant, impairment in recognition for low accessibility items related to those participants practiced fabricating about, and the Bayesian analyses again suggested that this may be due to low statistical power.

The impairment in recall was unique to lying participants. Experiments 1a and 2 failed to replicate the standard RIF effect in truth tellers. Despite using a variety of stimuli and switching to a between-subjects design so that truth telling was procedurally identical to a standard RIF experiment, selective truthful retrieval practice did not affect the accessibility of related information in memory. The standard RIF effect was found only in Experiment 1b, but this is likely to be due to output interference during recall, rather than genuine forgetting caused by an inhibitory mechanism engaged during retrieval practice.

Overall, these experiments suggest that repeating fabrications has no effect on memory for the corresponding truthful information but may impair memory for related information via an inhibitory mechanism. The RIF effects found have been discussed in Section 2.13, and I therefore consider in more detail here what might explain the null effects.

2.18.2 Why was there no RIF of corresponding truthful information?

Repeating fabrications did not impair recall below baseline for the corresponding truthful information in any experiment. This finding indicates that, in this context, practicing lies was equivalent to not rehearsing the items in the interval between learning and recall. This is a surprising finding, as these items were cued in this interval—they were directly probed in the retrieval practice phase. Despite being cued, there were no detrimental effects on memory (and even a slight facilitative effect for high accessibility items in Experiment 3). This suggests that the activation-competition-inhibition cycle outlined in Section 2.1.2 that induced RIF was not initiated when participants repeated their fabrications.

As outlined in Section 2.1.2, research into the cognition of deception indicates that the three main processes that induce RIF are also recruited when people fabricate—activation, competition, and inhibition. Although the present experiments do not provide direct evidence that these processes were not engaged when people repeated their fabrications, they suggest that at least one of the three components that induce RIF was not engaged. One possibility is that the retrieval cues provided in the present experiments did not activate the truth to the extent that it interfered with the generation of a deceptive response. This raises some interesting theoretical questions regarding truth automaticity in deceptive responding.

It is generally agreed among deception researchers that one of the first cognitive steps involved in deception is automatic activation of the truth in working memory (Debey et al., 2014; Walczyk et al., 2014). Truth automaticity depends first and foremost on its accessibility; if the truth is not accessible, it cannot automatically be retrieved. But even when the truth is accessible, there may be other influences on the extent to which it is activated and therefore the cognitive processes that follow. One such influence might be the way in which participants lie. Much of the research into the cognition of deception requires participants to lie by providing the opposite answer to the truth in response to yes/no questions (Debey et al., 2014; Duran et al., 2010; Suchotzki et al., 2017). This procedure is typically used because it allows a high degree of experimental control and for specific inferences to be made about the cognitive processes recruited by measuring participants' RTs under different conditions. Such research has shown that lying participants take longer to respond than truth telling participants—a robust effect called *the lying constant* (Sheridan & Flowers, 2010). The additional time to respond is taken as evidence that liars must override the automatic tendency to respond truthfully in order to lie (Duran et al., 2010; Williams et al., 2013).

However, lying by providing the opposite response to the truth is a very constrained form of lying. In such cases, strong activation of the truth might play an important role because the probability of inadvertently answering truthfully is 50% (given that there are two possible answers: yes or no). With just two possible response options, it is important that the truth is first known so that the opposite answer can be delivered. However, activation of the truth may play a less important role when deception is unconstrained—as in the present experiments where participants were free to fabricate whichever response they liked. In these cases, the probability of inadvertently answering truthfully is much lower, given that any answer can be provided, so long as it is accessible and plausible. Fabrication therefore does not require such strict monitoring of the truth. This idea is consistent with past research showing that people more often report needing to suppress the truth when lying in response to closed, rather than open-ended questions (35.5% and 82% of cases for open- and closed-questions respectively; Walczyk et al., 2003).

It is therefore possible that the activation, competition, inhibition cycle that leads to RIF may only be apparent in cases of constrained deception. Past research has studied something similar: Whether practicing lies by responding with the opposite to the truth reduces the cognitive cost of lying (Van Bockstaele et al., 2012). Practicing lies in this way decreased participants' RTs for the items they lied about. One explanation for this finding may be that repeated inhibition of the truthful response during practice resulted in less interference from the truthful response later on. Thus, one avenue for further research is to combine the RIF paradigm with a speeded constrained deception paradigm to determine if lying in this way can impair memory for the truth.

2.18.3 Theoretical and practical implications

As discussed in Section 1.4, several researchers have proposed that RIF may be a mechanism underlying the deception-induced memory impairments found in prior research. Von Hippel and Trivers (2011) also propose that by repeating lies and omitting the truth, people can come to forget the truth via the process of RIF, and therefore self-deceive. To conclude that RIF is an underlying mechanism of any forgetting effect, memory must fall below the individual's baseline forgetting rate and should also reflect a genuine memory impairment, rather than an artifact of the way that memory is tested.

The experiments described in this chapter are the first the use a paradigm that is designed to elicit RIF to test the possibility that the accessibility of the truth can be reduced merely by repeating fabrications.

If the purpose of self-deception is to help the liar to escape detection, then any memory impairment should be for the corresponding truthful information, since it is this information that the liar wishes to conceal. However, there was no evidence that repeating fabrications and omitting the truth can impair memory for the corresponding truthful information in any of the experiments reported here. Thus, these experiments do not provide evidence that RIF may be a mechanism underlying self-deception.

Nonetheless, repeating fabrications did impair memory below baseline for related truthful information. Memory can therefore suffer merely from repeating fabrications and omitting the truth (although it should be noted that the impairment was small). In a criminal investigation, this could mean that crime-relevant information that is not the subject of initial investigations can become less accessible as a result of fabricating false information about other crime-relevant details. If such information later becomes relevant to the investigation, the individual may find it easier to lie about these details and be more likely to escape detection as a result. In terms of the practical implications of such forgetting, a question for further research is whether the forgetting effects persist.

In sum, the primary purpose of the experiments reported here was to test the proposal that RIF is a possible mechanism underlying self-deception. No evidence for this claim was found. It remains possible that people can come to forget the truth or believe their lies after having repeated them and omitting the truth, however, there is currently no evidence that RIF underlies this effect. I therefore move onto the second mechanism by which people could come to believe their own lies: Source monitoring, and in particular, its relationship with cognitive load.

Chapter 3: Does the cognitive effort of lying accurately predict subsequent memory errors? Correlation, causation and Simpson's Paradox in the memory and deception framework

As outlined in Section 1.5.1, there is considerable existing evidence that people can come to mistake their lies for the truth via source monitoring errors and incipient research investigating the factors that moderate such errors. One potential moderator is the cognitive load that a liar experiences in the generation and delivery of the lie. Indeed, the recently developed MAD framework goes one step further by proposing that cognitive load might directly cause source monitoring errors, rather than merely moderate those errors. In this chapter, the hypothesis that liars' cognitive load causes subsequent memory errors is questioned. It is also argued that cognitive load might not be a useful predictor of the memory errors that follow deception. Chapter 4 then tests the claims made in the present chapter.

3.1 The MAD framework revisited

Since the MAD framework (Otgaar & Baker, 2018) is detailed in Section 1.6, only a summary is provided here. To recap, the MAD framework is based on the observation that the type of memory error that follows deception seems to depend on the liars' cognitive load. Specifically, lies that require fewer cognitive resources to generate (false denials and feigned amnesia) typically increase omission errors, whereas lies that require more cognitive resources to generate (fabrications) typically increase commission errors. The central hypothesis of the framework is therefore that low cognitive load lies should increase omission errors (forgetting of the truth), whereas high cognitive load lies should increase commission errors (source monitoring errors or false memories). Moreover, the framework proposes that increasing the cognitive effort required to lie should increase commission errors. In what follows, I question the validity of these hypotheses.

3.2. Cognitive effort is a marker, but not a mechanism

Although cognitive effort is the central concept underpinning the MAD framework, it is unclear how the cognitive effort of lying helps to explain subsequent memory errors. Despite stating that the effect of lying on memory is the result of the cognitive effort required, the MAD framework does not treat cognitive effort as a mechanism that *causes* omission and commission errors, but rather treats cognitive effort as a marker that *correlates* with other processes that cause omission and commission errors.

The MAD framework does not attribute the increased omission errors resulting from false denials and feigned amnesia to low cognitive effort, but instead to a lack of rehearsal. Specifically, the framework states that "when the draw of cognitive resources is less (such as in simple versions of false denials), issues such as a lack of rehearsal come into play and lead to omissions" (Otgaar & Baker, 2018, p. 9). In this statement, the fact that these lies require fewer cognitive resources is descriptive, but not explanatory. Instead, a lack of rehearsal is stated as the mechanism that explains the omission errors resulting from false denials and feigned amnesia.

Similarly, in the case of fabrication, increased commission errors are not attributed to high cognitive effort, but to source monitoring difficulties. Specifically, the framework states that "when more cognitive resources are required, participants must remember exactly what they lied about and remain consistent. As a result, the threat of source monitoring errors is lurking, which potentially leads to commission errors/false memories" (Otgaar & Baker, 2018, p. 9). Again, the fact that these lies require greater cognitive resources is descriptive, and source monitoring instead is the mechanism that explains the commission errors resulting from fabrication.

These mechanisms—a lack of rehearsal and source monitoring—do the heavy lifting in explaining the different memory errors produced by different types of lie. Cognitive effort, however, adds little to our understanding of how lying affects memory over and above the mechanisms with which it is associated. And that is because it is not obvious how cognitive effort should affect omission or commission errors without reference to these other causal mechanisms.

Why does it matter if the cognitive effort of lying correlates with other processes, but does not itself cause the different memory errors? After all, even if cognitive effort plays no causal role in producing memory errors, it may still legitimately predict the memorial outcome of lying, as long as it correlates highly with the actual causal processes. To draw an analogy, it might be possible to predict the number of shark attacks in a season from the water temperature, despite the fact that shark attacks are not caused by warmer water, but rather by an increase in the number of swimmers on warmer days. The problem is that the water temperature can dissociate from the number of swimmers; that is, there may be days when there are many swimmers, but cold water, and vice versa. On these days, your prediction will be inaccurate if based on the water temperature alone. Thus, if you know that shark attacks are caused by more swimmers, not by warmer water, then it is preferable to base your prediction on the number of swimmers accordingly to strive for maximal accuracy. In general terms, robust theories or frameworks center on causal mechanisms, not on correlated variables.

Returning to the MAD framework, the key issue can be summarized as this: MAD treats the cognitive effort of lying as analogous to water temperature in the above example (i.e. a correlate), but simultaneously encourages researchers to use it as a predictor, as if it were analogous to the number of swimmers (i.e. a cause). But the cognitive effort of lying and the type of lie are not interchangeable predictors: The cognitive effort of lying can dissociate from the type of lie in the same way that water temperature can dissociate from the number of swimmers. To illustrate, fabrication typically requires greater cognitive resources than does false denial or feigning amnesia, but not necessarily so. For instance, an individual guilty of committing some crime might find it very easy to fabricate a cover story when an investigator has little evidence or limited knowledge of the crime, since the individual does not have to worry about their story being consistent with the investigator's knowledge. Moreover, this individual might expend less cognitive effort to fabricate a story than somebody who instead falsely denies information in response to a knowledgeable investigator and therefore must think carefully about what they can deny, given what the investigator knows or what they themselves have disclosed in the past. Thus, there might be cases where fabrication is easier than selectively denying information. The cognitive effort required within a single lie type can also significantly vary depending on situational and individual factors. For instance, a well-practiced liar might require few cognitive resources to fabricate a lie, but a primarily honest individual might require many. Similarly, a more elaborate fabrication will likely require many more cognitive resources than a simple one.

In cases like those above where the type of lie dissociates from the cognitive effort required to create it, it is unclear what type of memory error the MAD framework would predict. For instance, if a fabrication requires few cognitive resources, should we expect increased commission errors because fabricating information increases vulnerability to source monitoring errors, or should we expect increased omission errors because the lie required little cognitive effort? This uncertainty stems from predicting memory errors from a non-causal variable. If we are to maximize the accuracy of our predictions, we will need to predict subsequent memory errors from their underlying mechanisms. Any theory of how lying affects memory should therefore center on these underlying mechanisms.

In the next section, it is explained why failing to dissociate two distinct variables—in this case, the cognitive effort of lying and the type of lie—can lead to incorrect interpretations of the existing empirical evidence. Specifically, pooling research on the basis of a non-causal variable can lead to Simpson's Paradox and consequently to incorrect statistical inferences.

3.3 The cognitive effort of lying and commission errors: A potential case of Simpson's Paradox

Simpson's paradox occurs when a relationship between variables reverses direction when controlling for additional variables. It is best illustrated with an example. In 1973, the University of California-Berkeley was sued for an apparent gender bias in graduate admissions. Overall, 44% of male applicants were admitted, compared to only 35% of female applicants, showing a clear bias in favor of admitting men. However, when the academic department was factored into the analysis, the pattern reversed, showing a small bias in favor of admitting women (Bickel, Hammel, & O'Connell, 1975). Women more often applied to Humanities departments that had few places and were therefore difficult to get accepted into, whereas men more often applied to mathematical and science departments with more places. Differences in the number of places available across departments helped to explain the differing admission rates between genders, but this was ignored in the initial analysis—it was a *lurking explanatory variable*. Instead, the admission data were improperly pooled into a “gender” variable, resulting in a spurious relationship between gender and admission rates. How does this relate to the MAD framework? Focusing on commission errors, the next section shows how pooling research into a cognitive effort continuum, as the MAD framework does, could lead to an instance of Simpson’s Paradox.

Because existing research on the effect of lying on memory has not explicitly manipulated the cognitive effort of lying, we can currently only infer the relationship between the cognitive effort of lying and commission errors. Instead, the research to date has studied the effects of different types of lie on memory. From this, we know that false denials and feigned amnesia (which typically fall at the low end of the cognitive effort continuum) produce few commission errors (but more omission errors), but fabrications (which typically fall at the high end of the continuum) produce more commission errors. Thus, plotting the rate of commission errors against the cognitive effort of lying might result in something that approximates Figure 3.1, in its simplest form.

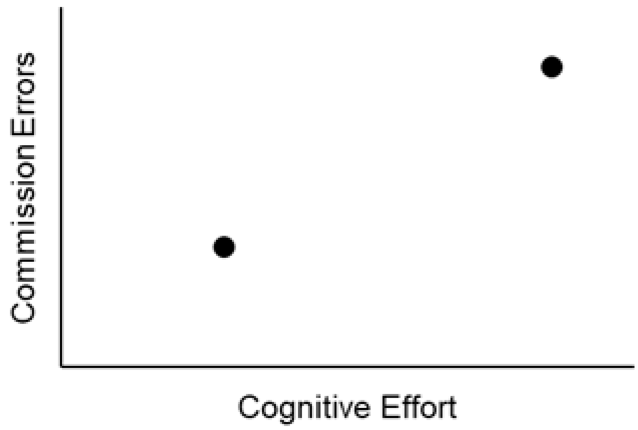


Figure 3.1. Schematic diagram plotting the rate of commission errors for lies at either end of the cognitive effort continuum.

It is clear from the figure that we know little about the commission errors produced from lies in the middle of the cognitive effort continuum—lies that are more cognitively demanding than denial or feigning amnesia typically is, but less cognitively demanding than fabrication typically is. As a result, we do not have enough information to join the data points in Figure 3.1. That is, we do not know if the relationship between commission errors and the cognitive effort of lying is linear (commission errors increase proportionally with cognitive effort) or nonlinear (commission errors increase disproportionately with cognitive effort). The MAD framework states that "the side effect of increasing cognitive resources during deceptive attempts such as fabrication is that lying will have similar characteristics as genuine memories" (Otgaar & Baker, 2018, p. 8). Thus, by pooling existing research into a single cognitive effort continuum, the framework implies a positive association between the cognitive effort of lying and the rate of commission errors.

Pooling research findings can be problematic when the variable into which they are pooled is confounded with other variables, as in this case. There is currently no research that dissociates the cognitive effort of lying from the type of lie participants are instructed to tell. As such, the conclusions about the effect of cognitively demanding lies on memory have been drawn from research into one type of lie (i.e., fabrication), but our conclusions about the effect of cognitively less demanding lies on memory have been drawn from research into a different type of lie (i.e., denial, feigned amnesia). In other words, it is not yet possible to make any claims about the relationship between the cognitive effort of lying and the rate of commission errors because we cannot separate the effects of cognitive effort from the effects of the type of lie that was told. Determining the role of cognitive effort requires manipulating the cognitive effort of the same type of lie and seeing what effect this has on the memory errors that follow (Experiments 4 and 5 described in Chapter 4 are designed for this purpose). Moreover, because the cognitive effort of lying does not cause commission errors per se (as discussed in Section 3.2), there is likely to be a lurking explanatory variable that explains the apparent increase in commission errors as we move up the cognitive effort continuum.

The lurking explanatory variable in question may simply be the type of lie told. As the MAD framework highlights, fabricating an event creates false details in memory that can be confused for the truth, whereas denying an event or feigning amnesia does not. This difference seems sufficient to explain why fabrication can create commission errors, without invoking cognitive effort. Thus, the increase in commission errors from lies at the high end of the cognitive effort continuum might be primarily explained by a shift in the type of lie, from those that do not create alternative imagined details to those that do. But to be an instance of Simpson's paradox, controlling for the type of lie must reverse the apparent positive relationship so it becomes negative, that is, commission errors should *decrease* as the cognitive effort of lying increases.

3.4 Why might controlling for the type of lie imply a different relationship between lying and commission errors?

The MAD framework cites the SMF (Johnson et al., 1993) to explain why lies requiring greater cognitive resources can increase commission errors. But the SMF can also predict the opposite result that lies requiring greater cognitive resources can *decrease* commission errors. As outlined in Section 1.5, the SMF predicts that internally generated information (such as fabricated information) is more likely to be incorporated into memory when it has characteristics that resemble a true memory. But recall that the SMF also posits another factor that influences whether fabricated information will be incorporated into memory: Cognitive operations. To recap, these are the mental processes that are activated when encoding information, for example, the retrieval processes engaged when generating fabricated information and the processes that led you to choose to report certain details over others (Foley & Foley, 2007).

According to the SMF, the cognitive operations associated with a memory help people to identify when information was generated internally versus based on a real experience, and consequently, commission errors are reduced when relatively more cognitive operations are associated with a memory (Johnson et al., 1988; Sugimori & Tanno, 2010). Thus, whereas fabricating information might produce characteristics typical of genuine memories, the act of fabrication might also produce cognitive operations that prevent this information from being incorporated into memory. Moreover, it may be that the more effortful—or cognitively demanding—the process of fabrication is, the more cognitive operations are associated with the fabricated information. Fabrications that require more cognitive resources might therefore result in *fewer* commission errors than do fabrications that require fewer cognitive resources.

There is also preliminary empirical evidence suggesting that increasing the cognitive effort of lying can decrease commission errors. For instance, studies comparing the effect of self- vs other-generated fabrications show that other-generated fabrications are more likely to increase commission errors, suggesting that the additional cognitive processes recruited to self-generate lies somewhat protected participants from incorporating those lies into memory (Pezdek et al., 2009; Pickel, 2004). Consistent with this line of reasoning, we know that frequent liars are more likely to incorporate fabricated details into memory than are those who lie less often (Polage, 2012). Practiced liars presumably find it easier to lie and therefore need to recruit fewer cognitive resources to do so. Nonetheless, because these studies did not directly manipulate cognitive effort, we cannot be sure that this explains the observed differences. However, they do suggest that a negative association may exist between the cognitive effort of fabrication and the rate of commission errors.

It is an empirical question whether increasing the cognitive effort of fabrication decreases or increases commission errors. But this can only be established when research dissociates the cognitive effort of lying from the type of lie. If future research does not do this, then it might seem that the association between the cognitive effort of lying and the rate of commission errors is positive, even if it is actually negative. In other words, a negative association could be masked by the fact that cognitive effort is confounded with lie type. The next section illustrates this.

3.5 Why might pooling commission errors on a cognitive effort continuum lead to interpretive issues?

Imagine that we conducted an experiment to see how lying affects commission errors at different points along the cognitive effort continuum, but we do not control for the type of lie told. In this imaginary experiment, we ask people to deny an event (as a low cognitive effort condition), fabricate a one-word answer (as a mid-cognitive effort condition), or fabricate an elaborate answer (as a high cognitive effort condition). Suppose that our hypothesis is that the rate of commission errors will be higher for lies requiring greater cognitive effort, in line with MAD, and that the results show something like the pattern in Figure 3.2.

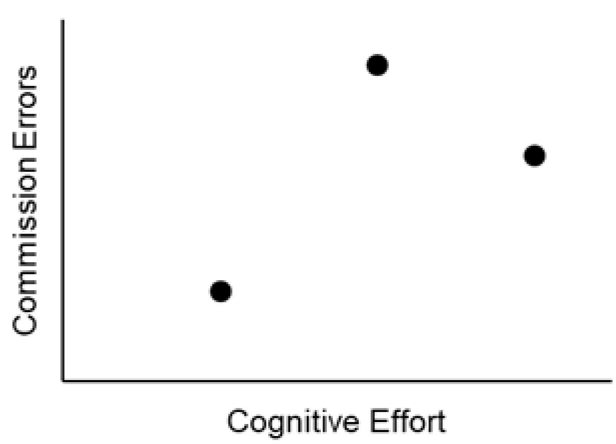


Figure 3.2. Schematic diagram depicting the results of an imaginary experiment that plots the rate of commission errors for lies requiring low, mid, and high cognitive effort.

If we modelled the data in Figure 3.2, with cognitive effort as our predictor, we would see a positive association between the cognitive effort of lying and commission errors, as depicted in the left panel of Figure 3.3. But this apparent positive association is misleading because it neglects the fact that the data points come from different types of lie—the lurking explanatory variable. Once we include the type of lie in the model, the association reverses, as depicted in the right panel of Figure 3.3. Thus, the apparent positive association between the cognitive effort of lying and commission errors might actually be a negative association once the type of lie is controlled for. In other words, there may be an interaction between the cognitive effort of lying and lie type, such that the effect of cognitive effort on commission errors depends on the type of lie told. But unless the research explicitly dissociates the cognitive effort of lying from the type of lie, we cannot express this relationship as an interaction and this can lead us to misinterpret the findings.

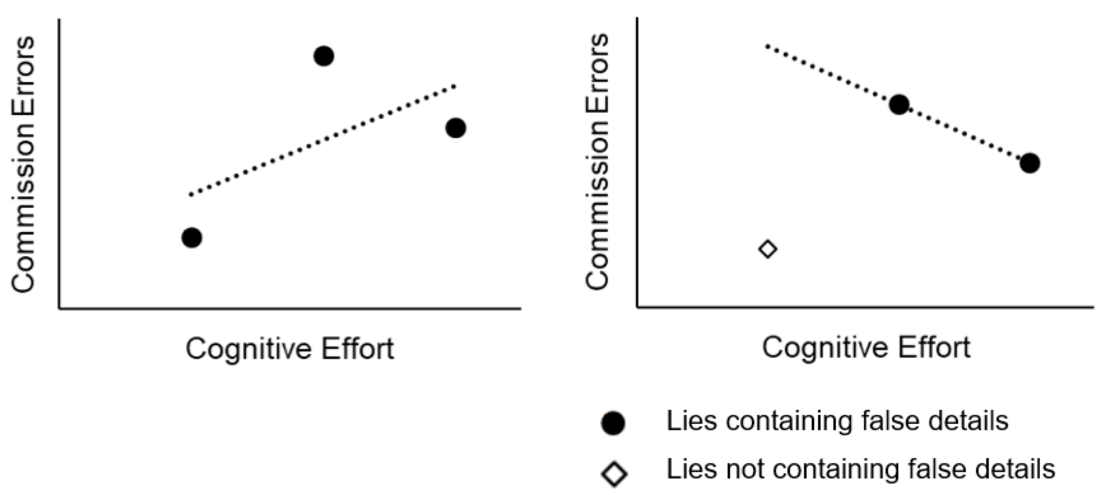


Figure 3.3. Schematic diagram depicting Simpson's paradox. Left panel: A spurious positive association between the cognitive effort of lying and commission errors when data are pooled along a single cognitive effort continuum. Right panel: A negative relationship between the cognitive effort of lying and commission errors when controlling for the type of lie.

3.6. Conclusion

By synthesizing the existing research on lying and memory, Otgaar and Baker (2018) have imposed a much-needed structure on this area of research. Despite many missing puzzle pieces, they developed a plausible and testable framework that can greatly further our understanding of how lying affects memory. Nonetheless, the framework confounds distinct variables, which could lead to interpretive issues.

This chapter began by stating that a good theory or framework predicts an outcome from a causal mechanism because this will yield the most accurate predictions. The MAD framework predicts the type of memory error that lying produces based on the cognitive effort of lying. However, it seems that the cognitive effort of lying does not cause memory errors, but instead correlates with other causal mechanisms (namely, a lack of rehearsal and source monitoring difficulties). Predicting the memorial outcome of lying from cognitive effort could therefore lead us to draw the wrong conclusions. This is because pooling lies, irrespective of the type of lie, into a single cognitive effort continuum might lead us to overlook the actual cause of the different memory errors. As we have seen, this improper pooling can lead to Simpson's paradox—what looks like a positive association between cognitive effort and a certain type of memory error might actually be a negative association when we control for lurking explanatory variables.

The next chapter outlines 2 experiments that were designed to explicitly dissociate the cognitive effort required for lie generation and the lie type. The experiments tested the hypothesis that when controlling for the type of lie, commission errors may actually be higher when participants lie under low cognitive load and lower when they lie under high cognitive load (contrary to the MAD framework's prediction).

Chapter 4: Does the cognitive effort of lying affect whether fabrications are incorporated into memory?

4.1 Introduction

The previous chapter questioned the MAD framework's assumption that the cognitive effort of lie generation and the type of lie told are interchangeable predictors of the memory errors that may follow deception. Instead, it was proposed that the cognitive effort of lie generation should be treated as a distinct variable that can moderate the effect of different types of lie on memory. In real life, there are several factors that may dissociate the cognitive effort of lying from the type of lie told. For instance:

- Practice—an individual who fabricates often will likely require fewer cognitive resources to generate their fabrication than somebody who fabricates rarely. This is supported by research showing that people are quicker at lying after practice (Van Bockstaele et al., 2012). Similarly, spontaneous fabrication generally imposes greater cognitive load than delivering pre-prepared fabrications (Walczyk et al., 2014).
- The length of the lie—single- or few-word fabrications will likely require fewer cognitive resources than lengthy narrative fabrications (Walczyk, Igou, Dixon, & Tcholakian, 2013)
- Plausibility—ADCAT (Walczyk et al., 2014) proposes that the cognitive cost of fabrication is inversely related to lie plausibility, such that constructing more plausible fabrications requires less cognitive effort. This is because plausible fabrications are typically closer to the truth and are therefore easier to construct. Less plausible fabrications tend to result from limited experience with the event in question and therefore tend to be constructed from more general schemata and scripts (Sporer, 2016), rather than specific experiences. This requires greater cognitive effort to the extent that it requires the manipulation of knowledge stores.
- Situational factors, such as the knowledge of the enquirer. If the enquirer has extensive knowledge of the event in question, it is likely that more cognitive

effort will be required to construct fabrications that are consistent with the known facts compared to when the enquirer knows little.

While all of the above may be natural variants of the cognitive effort required for fabrication, they cannot form the basis of an experimental manipulation to test the hypothesis that increasing the cognitive effort of lie generation decreases commission errors. This is because they all confound cognitive load with another factor. For instance, if cognitive load is manipulated by asking participants to tell spontaneous versus practiced lies, it cannot be known if any later differences in memory are due to the diminished cognitive load from practice, or because practice increases fluency and affects source monitoring in other ways.

To isolate the effects of cognitive load independently of its covariates, the following two experiments manipulated the cognitive effort of fabrication using anagrams. Participants were provided with a ready-made fabrication in the form of an anagram that first must be solved in order for the lie to be delivered. The difficulty of the anagram was used as a proxy for the effort required to construct the fabrication: The harder the anagram, the more the individual contributes to the lie's construction. This allowed control over the type of lie as well as the specific lie told. All participants reported the same fabrications, but had to put varying effort into the construction of that fabrication according to how difficult it was for them to solve the anagram. This provided a strictly controlled method of manipulating the cognitive effort required for lie construction. Anagrams are frequently used in broader memory literature to examine the relationship between cognitive effort, cognitive operations and subsequent memory performance (e.g., Bertsch, Pesta, Wiscott, & McDaniel, 2007; Foley & Foley, 2007; Taconnat & Isingrini, 2004). Although this method compromises external validity, it provides good experimental control to test the hypothesis that the cognitive effort of lie construction can dissociate from the type of lie told with different memorial outcomes.

4.2 Experiment 4

Experiment 4 tested the hypothesis developed in Chapter 3 that commission errors will be higher when the cognitive effort of lying is low and lower when the cognitive effort of lying is high—the opposite prediction to the MAD framework. Importantly, the type of lie was controlled such that all participants delivered the same fabrications. Since participants did not construct their own lies, this is analogous to when an individual reports details heard from someone else, or when an accomplice gives the individual their alibi.

The experiment was conducted over 2 sessions. In Session 1, participants watched an interactive mock crime before undergoing interview or filler tasks (a no fabrication control condition). All participants lied at interview under conditions of low or high cognitive load. Participants either simply received the lie to deliver (low load condition) or they received the fabrication as an anagram that required solving before the lie could be delivered (high load condition). Three weeks later, participants completed a recognition test to assess their memory for the items from the mock crime. There were two main predictions: First, that participants who provide false answers (fabrications) to questions about the mock crime will show more commission errors, poorer recognition discriminability, and lower confidence than participants who do not fabricate any details. Second, that contrary to the MAD framework's prediction, participants in the low load condition will show more commission errors, poorer recognition discriminability and lower confidence than participants in the high load condition. The study was pre-registered on the Open Science Framework at the following link: https://osf.io/y6es4/.

4.3 Method

4.3.1 Participants.

The experiment was completed by 159 participants (mean age = 19.6 years, *SD* = 2.5 years, range = 18-34 years; 38 male, 120 female, 1 other) in exchange for course credit or a small monetary reward. Participants were primarily native English speakers (n=119). Of those whose first language was not English, 31 rated themselves as fluent to a native or excellent standard and the remaining 9 subjects rated their English as either good (n=7) or satisfactory (n=2).

A power analysis in R was conducted to determine an appropriate sample size using the BUCSS package (Anderson & Kelley, 2019) that adjusts effect sizes from previous studies for publication bias and uncertainty (Anderson, Kelley, & Maxwell, 2017). The power analysis was based on Pezdek et al.'s (2009) forced fabrication study, as the design and study goals most closely approximated those of the present experiment. In Pezdek et al., some participants were forced to fabricate their own answers to questions concerning a mock crime they witnessed (self-generated), whereas other participants answered with a fabrication suggested in the question (other-generated). The power analysis was based on this between-subjects comparison, which is analogous to the high and low cognitive load comparison in the present experiment (self-generating a fabrication is likely to be more cognitively demanding than merely accepting one supplied in the question). Based on the adjusted effect size of d=0.79 from Pezdek et al., the present experiment required 51 participants per group (a total of 153 participants) to achieve a power of 0.8 with an alpha level of .05 and assurance of 0.7 (assurance is the percentage of times that the experiment would reach the specified power (Anderson et al., 2017). The study was approved by the psychology department ethics committee at the University of Warwick.

4.3.2 Stimuli and apparatus.

Mock crime. This experiment used the mock crime video that was used in Experiment 2 (only the crime-relevant video was used in the present experiment). To recap, the video was filmed from a POV perspective and depicted a gloved individual carrying out numerous actions culminating in the theft of an expensive device and some important documentation. For the present experiment, the mock crime video was edited to make it interactive. At various points throughout the video, participants were required to select the correct item to carry out an action described (this is outlined in more detail in Section 4.3.3). This was to make the mock crime more engaging and ensure that participants remained focused throughout the video, ultimately to promote better encoding. Participants watched the video full screen on 22-inch monitors with a resolution of 1920x1080 (duration = 8 minutes, 50 seconds). The interactive mock crime was developed using PsychoPy (Peirce, 2007).

Lie creation. Participants were provided with the fabrications to deliver during the interview. A pilot survey was conducted to ensure that the fabrications were plausible alternatives to the truthful corresponding details from the mock crime (see Appendix 8).

In the high cognitive load condition, the fabrication was presented as an anagram that required solving before the lie could be delivered. The anagrams were developed using the same method as past research (Foley & Foley, 2007). A pilot study was conducted to ensure that participants could solve the anagrams, but also considered them more cognitively demanding to solve than merely reading the answer (see Appendix 8).

4.3.3 Design and Procedure.

The experiment was fully between-subjects and included a single independent variable with 3 levels (Fabrication Type: low cognitive load, high cognitive load, no fabrication/control). The experiment involved two sessions, completed 3 weeks apart. Session 1 consisted of a perpetration, distraction and interview phase, and was completed in individual laboratory cubicles. Session 2 was a recognition test that participants completed online. Participants were given a cover story to prevent them from ascertaining the study's true purpose– they were informed that the experiment concerned verbal lie detection over time (a suspicion check was included to determine the success of the cover story; the results are reported below).

Session 1

Mock crime phase. After consenting to participate and providing demographic information, participants read a background story to immerse them in the mock crime scenario. They were asked to imagine that they had been in financial difficulty and resorted to borrowing money from a loan shark, who they were unable to repay. In a mock phone call, the angry loan shark informed the participant that they could wipe their debt if they break into a university building to steal some valuable equipment and documentation. Participants then received instructions on how to execute the crime to clear their debt. The instructions consisted of 18 statements, each outlining an action together with the item needed to complete the action written in red (e.g., "lever open the workshop door using a crowbar"). Participants were instructed to memorize the items and had 3 minutes to do so. The first and last items were filler items to control for primacy and recency effects, leaving 16 critical items for memorization.

Participants ‘performed’ the mock crime by watching the interactive mock crime video, consisting of the 18 actions outlined in the instructions. Before each action in the video was executed, an inventory appeared on the screen showing the 18 items and a description of the upcoming action (see Figure 4.1). Participants then selected the item required to complete the action from the inventory. The inventory items were arranged in a random order, but the order was kept constant across all 18 actions.

Distractor phase. Participants completed 10 minutes of mathematical, reasoning and word puzzles. The experiment automatically progressed to the interview phase after 10 minutes. More than 10 minutes-worth of tasks were included and no participants completed the given tasks in this time, suggesting that they were occupied for the full duration.

Interview phase. The computer program randomly allocated participants to either the high cognitive load, low cognitive load or no fabrication/control condition. Participants in the no fabrication/control condition continued to complete word puzzles for a further 10 minutes to prevent them from rehearsing the details of the mock crime and were then thanked and dismissed.

Participants in the low and high cognitive load conditions underwent interview on the actions carried out in the mock crime. The interview consisted of pre-recorded videos of a ‘rival’ of the loan shark, who was allegedly planning a similar crime and wanted to know how it was done. Participants were told to conceal their methods by lying in response to his questions by delivering the lie provided onscreen as convincingly as possible. The interview consisted of 18 subtitled videos, each containing a question about an item from the mock crime (e.g. “What did you use to cover the CCTV camera?”). The question videos played one-by-one in a random order.

For each video, once it finished playing, the lie appeared below the video (the stimuli are listed in Appendix 6). Participants in the low cognitive load condition simply read the lie, retyped it and stated it aloud (e.g. A CLOTH). Participants in the high cognitive load condition received the lie in the anagram form (e.g., "A COHTL") and were instructed to type the solution and state it aloud. If participants could not solve the anagram, a hint appeared on screen, consisting of an easier anagram that was solved by rearranging 3 underlined letters. To encourage participants to solve the anagrams themselves and not simply wait for the easier version, the hints appeared only after 90 seconds. In both conditions, participants were reminded to be as convincing as possible when delivering their lies and instructed to picture the deceptive answer in their mind's eye. All deceptive answers were one or two words in length. The first 2 questions were practice questions addressing the filler items to familiarize participants with the task. Feedback was provided for these questions.

On completion of the interview, participants completed a manipulation check to determine whether the cognitive load manipulation worked. Participants rated how much mental effort they required to answer the interview questions on a scale of 1-9 ranging from "very low mental effort" to "very high mental effort". Past research has shown this to be a reliable index of cognitive load (Paas, Van Merriënboer, & Adam, 1994). This concluded Session 1.

Figure 4.1. The item inventory for the interactive mock crime. Participants selected the correct item to carry out the action described for each of the 18 actions in the mock crime video.

Session 2

Participants were emailed a link to complete Session 2 of the study online 21 days after completing Session 1. They were directed to a yes/no recognition test that tested their memory for the 16 critical items from the mock crime. The instruction was to "answer all questions truthfully, i.e., "yes" if you saw the item in the mock crime video and "no" if you did not see the item in the mock crime video, regardless of whether you lied in response to these questions in part 1". Participants completed a comprehension check to ensure that they understood this instruction. There were 2 questions for each item (yielding 32 questions in total) – one containing the truthful answer, requiring a "yes"' response (e.g. "was the CCTV camera covered with shaving foam?") and one containing the deceptive answer, requiring a "no" response (e.g. "was the CCTV camera covered with a cloth?"). Participants indicated "yes" or "no" and rated their confidence that their answer was correct on a scale of 0 = *completely uncertain* to 100 = *completely certain*. Each question appeared one-by-one in the center of the screen and the question order was randomized across participants. The first 2 questions were practice questions addressing the filler items and participants were given feedback for these questions. On completion of the recognition test, participants were asked to state what they believe the study's hypothesis is, as a suspicion check. No participants correctly guessed the study's hypothesis.

4.4 Results and discussion

4.4.1 Preliminary analyses

Before addressing the primary research question of the experiment, four manipulation and performance checks were performed.

Engagement in mock crime. Participants performed well in the interactive mock crime and selected the correct item to perform each action for most of the instructions (M = 94.58%, SD = 22.64%). This indicates that participants were engaged with the mock crime scenario and encoded the crime-relevant items well.

Manipulation check. Participants' ratings of the mental effort required to complete the interview suggest that the cognitive load manipulation was unsuccessful: The difficulty ratings for the low load and high load conditions did not significantly differ ($M_{\text{low load}}$ = 3.30, SD = 1.85; $M_{\text{high load}}$ = 3.72; SD = 1.75; t(104)= -1.19, p = .23). Participants did, however, take significantly longer to answer each interview question in the high cognitive load condition (M = 19.56s, SD = 14.84s) than the low cognitive load condition (M = 15.87s, SD = 6.7s; (t(104) = 4.23, p < .001, 95%CI[1.95, 5.39]. This suggests that the anagram solutions were not immediately obvious and therefore that participants did have to think harder to solve them than to merely read the word in the low load condition. It is therefore plausible that participants in the high load condition still required greater cognitive effort to answer the interview questions than participants in the low load condition, but this difference is nonetheless smaller than anticipated.

Interview performance. First, participants' verbal responses to each interview question were transcribed to ensure that they both typed and stated aloud the correct answer to each. Participants in the low load condition were merely retyping and saying the lie written onscreen and therefore achieved perfect performance for typing their responses

(*M* = 100%, *SD* = 0%) and almost perfect performance for stating their responses aloud (*M* = 99.88%, *SD* = 3.43%). Participants in the high load condition typed and stated aloud the correct anagram solution for most questions (M_{typed} = 97.05%, *SD* = 16.93%; M_{aloud} = 97.41%, *SD* = 15.91%). If participants in the high load condition did not solve the anagram within 90 seconds, they received a hint in the form of an easier anagram. The hint was required for 1.15% of questions (collapsed across all participants).

Attrition across sessions. Two participants failed to complete Session 2. Complete recognition data was therefore obtained from 52 participants in the control condition, 52 participants in the high load condition and 53 participants in the low load condition.

The average amount of time elapsed between completion of Session 1 and 2 was 21.50 days (*SD* = 0.73 days) and the time between sessions did not differ significantly across conditions, $F(2, 156) = .06$, $p = .95$.

4.4.2 Signal detection theory and analysis

Signal detection measures were used to assess recognition performance. Signal detection theory (SDT) can be applied whenever an individual must distinguish a signal (previously encountered stimuli) from noise (new stimuli; Stanislaw & Todorov, 1999). Signal detection measures are therefore commonly used in yes/no recognition tests that include signal and noise trials, as in the present study. Signal trials were questions containing truth items, that is, items that were experienced in the mock crime. Noise trials were questions containing lie items, that is, items that participants provided as a lie during interview.

An individual's performance in the recognition test can be defined by their hit rate (HR) and false alarm rate (FAR). The HR is the probability of correctly responding "yes" in truth trials (questions containing an item from the mock crime) and the FAR is the probability of incorrectly responding "yes" in lie trials (questions containing an item that participants provided as a lie). According to SDT, the HR and FAR do not only reflect people's recognition memory (their ability to distinguish lies from truths in memory, i.e., "discriminability"), but also their general tendency to answer either "yes" or "no", i.e., "response bias"). Since we are interested in whether lying affects recognition memory, a pure measure of discriminability is required that is not confounded with response bias. Without such a measure, it cannot be known whether lying alters people's ability to distinguish truths from lies in memory, or if lying merely alters people's willingness to state that an item was experienced or a lie (or both). Signal detection measures provide separate indices for an individual's discriminability and response bias to more precisely determine how lying affects later memory for the truth. *A*' (Pollack & Norman, 1964) and B" (Grier, 1971) were used as non-parametric measures of discriminability and response bias respectively (as recommended by Stanislaw & Todorov, 1999).

4.4.3 Recognition performance

To assess participants' performance in the recognition test, HRs (number of "yes" responses in truth trials / total number of truth trials) and FARs (number of "yes" responses to lie trials / total number of lie trials) were calculated for each participant. Table 4.1 shows the descriptive statistics for each condition.

FARs (commission errors) and HRs (omission errors). It was predicted that participants who lied during interview would make more commission errors (i.e., have a higher false alarm rate) in the recognition test than control participants who did not undergo interview. This hypothesis was supported: FARs differed significantly across conditions, as indicated by a between-subjects one-way ANOVA ($F(2,154) = 7.39$, $p < .001$). Furthermore, planned contrasts indicated that participants who lied at interview showed a significantly higher FAR than control participants ($M_{lie} = 0.23$, $SD = 0.12$, $M_{control} = 0.16$, $SD = 0.09$; $t(154) = 3.79$, $p < .001$, $d = 0.63$). Lying therefore increased commission errors, consistent with past research (e.g., Pickel, 2004; Rindal, 2017).

The HR also significantly differed across conditions (*Welch's* F[4](2, 100.41) = 3.0, $p = 0.05$). Participants who lied at interview showed a significantly lower HR than control subjects ($M_{lie} = 0.86$, $SD = 0.10$, $M_{control} = 0.90$ $SD = 0.07$; $t(154) = -2.14$, $p = 0.03$, $d =$

[4] Welch's F is reported due to unequal variances (as assessed by Levene's test; $F(2, 154) = 3.89$, $p = 0.02$).

-0.44), as indicated by planned contrasts. Thus, lying at interview increased both commission and omission errors: Participants who lied were more likely to mistake a lie for the truth and to remember fewer items from the mock crime than if they had not lied.

Both the FAR and HR were unaffected by the cognitive effort of lying. Although participants in the low load condition did indeed show higher FARs than participants in the high cognitive load condition (see Table 4.1), planned contrasts indicated that this difference was not significant ($t(154) = 0.65$, $p = 0.52$). Similarly, participants in the low and high load conditions did not significantly differ in their HRs ($t(154) = 0.89$, $p = 0.37$). The prediction that participants in the low load condition would make more commission errors (increased FARs) than participants in the high load condition was therefore not supported.

A' (discriminability) and B" (response bias). For the signal detection analysis, the formulae described by Stanislaw and Todorov (1999) were used to calculate each participant's A' (discriminability) and B" (response bias) values. The formula straightforwardly applied for A', but it was necessary to adjust the HRs and FARs for B" to accommodate a participant who achieved perfect performance (HR=1, FAR=0). B" values were adjusted using the loglinear approach (Hautus, 1995), which adds 0.5 to both the number of hits and false alarms and adds 1 to both the number of signal trials (questions containing an item from the mock crime) and noise trials (questions containing a corresponding lie item) for all participants. This means that the entire dataset is treated equally, whereas the commonly used approach to adjust signal detection values (Macmillan & Kaplan, 1985) only adjusts extreme values in the dataset. The loglinear approach therefore produces less biased values than Macmillan and Kaplan's approach (Hautus, 1995). Table 4.1 therefore shows the B" values based on the adjusted HRs and FARs.

As a measure of discriminability, A' typically ranges from 0.5 (indicating chance performance) to 1 (indicating perfect performance). We can see from Table 4.2 that discriminability was generally high, but participants who lied showed slightly lower discriminability than controls. A between-subjects one-way ANOVA revealed that A' values significantly differed across groups (*Welch's* F[5](2, 94.27) = 9.52, $p < .001$). As predicted, planned contrasts confirmed that lying significantly reduced participants' ability to discriminate items that they had stated as a lie from items that they experienced in the mock crime, compared to control participants (M_{lie} = 0.88, *SD* = 0.08, $M_{control}$ = 0.92, *SD* = 0.04; t(154) = -3.63, $p < .001$, $d = -0.58$). However, discriminability did not significantly differ between the low and high cognitive load conditions ($t(154) = 0.43$, $p = .67$). Thus, contrary to the hypothesis, the cognitive effort of lying did not affect discriminability.

Analysis of B" values was exploratory, as detailed in the study preregistration, because there is no precedent to hypothesize how varying interviewees' cognitive load while lying might affect response bias. B" values range from -1 (extreme liberal bias in favor of stating that an item was shown in the mock crime) to 1 (extreme conservative bias in favor of stating that an item was not shown in the mock crime), and a B" of 0 represents no bias. We can see from Table 4.1 that participants showed a slight liberal bias in all conditions (i.e., a general tendency to state that they experienced the item in the mock crime, regardless of whether they remember doing so.) Although participants in the low load condition showed a slightly stronger tendency to state that they experienced an item in the mock crime than participants in the control and high load conditions, B" values did not significantly differ across conditions ($F(2,154) = 1.69$, $p = 0.19$). Response bias was therefore unaffected by lying. This suggests that the increase in commission errors for participants who lied during interview is indeed attributable to poorer discriminability, rather than changes in response bias.

[5] Welch's F is reported, as A' variances were unequal (as assessed by Levene's test; F(2, 154) = 7.32, $p < .001$).

Table 4.1

Mean hit rate (HR), false alarm rate (FAR), discriminability (A'), and response bias (B") for each condition.

	HR (SD)	FAR (SD)	A'(SD)	B"(SD)
Control	0.90 (0.07)	0.16 (0.09)	0.92 (0.04)	-0.15 (0.32)
Low Load Lie	0.87 (0.10)	0.24 (0.11)	0.89 (0.07)	-0.23 (0.28)
High Load Lie	0.86 (0.10)	0.22 (0.13)	0.88 (0.08)	-0.14 (0.22)

4.4.4 Confidence

Participants' confidence in the accuracy of their memories was analyzed to determine if lying under varying levels of cognitive load leads to more subtle changes in memory than increased commission or omission errors. Participants rated their confidence that they had answered correctly for each question in the recognition test on a scale of 0-100. A new factor was created (*Accuracy)* by splitting the confidence ratings into 2 levels: ratings for questions participants answered correctly and ratings for questions participants answered incorrectly. The analysis focused on comparing confidence for questions containing lie items (fabrications) with questions containing truth items (items from the mock crime) across the three experimental conditions. The analysis plan in the study pre-registration stated that a one-way ANOVA would be performed to determine if confidence differs across the 3 conditions. However, it was likely that the item type and the accuracy of participants' responses would significantly affect the relationship between confidence and condition. Because the Accuracy and Item Type variables were within-subjects and because the levels of the Accuracy variable were unbalanced, the confidence data were analyzed using a multilevel linear model (see Section 2.16.2 for an explanation of multilevel linear models).

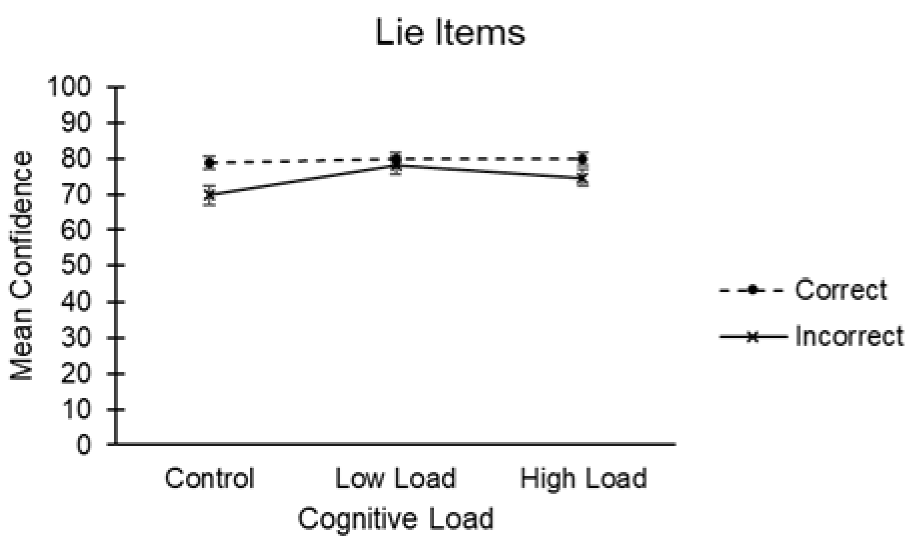
Model fitting. First, the fitting of the model to the confidence data is described and the model results are then interpreted. Figure 4.2 shows that confidence was generally high, but was clearly affected by participants' accuracy and whether the item in question was from the mock crime (truth item) or a fabrication (lie item). Indeed, including Accuracy as a predictor significantly improved the multilevel model fit compared to an intercept-only model ($\chi(6) = 140.61$, $p <. 001$), and adding Veracity further improved the model fit ($\chi(7) = 4.61$, $p = .03$). The cognitive load condition did not improve the model fit ($\chi(9) = 1.84$, $p = .40$). Adding the interaction term between Accuracy and Veracity further improved the model fit ($\chi(10) = 76.00$, $p <. 001$), however no other interaction terms significantly improved the model (all *p*s > .05). All non-significant variables and interaction terms were removed from the model. The final model therefore included the main effects of Accuracy and Veracity and the interaction between Accuracy and Veracity ($\chi(8) = 75.96$, $p <. 001$).

Model Interpretation. From Figure 4.2, we can see that confidence for questions containing truth items generally corresponded with accuracy, such that confidence was high when participants correctly stated that the item was from the mock crime ($M = 85.18$, $SD = 11.54$), but lower when they incorrectly stated that the item was not from the mock crime; that is, when they made an omission error ($M = 62.62$, $SD = 22.89$).

The difference in confidence for correct and incorrect answers was far smaller for questions containing lie items. Interestingly, confidence did not correspond with accuracy for lie items. Again, confidence was high when participants correctly stated that the item was not experienced in the mock crime ($M = 79.31$, $SD = 13.37$), but dropped very little when participants incorrectly stated that the lie item was experienced in the mock crime; that is, when they made a commission error ($M = 73.81$, $SD = 17.85$). Thus, when participants made a commission error, they were generally very confident that the lie was indeed the truthful answer. The model contrasts confirmed that the decrease in confidence for incorrect compared to correct answers was significantly greater when participants were asked about items from the mock crime (truth items) than when they were asked about the lies they told (lie items; $b = 17.43$, $t(281) = 9.08$, $p < .001$). Thus, participants confidence ratings aligned more with their accuracy for when they were asked about items from the mock crime compared to when they were asked about the lies they told.

Although participants' confidence in their memories was high when they made commission errors, the non-significant improvement to the multilevel model when including the cognitive load condition or any of its interaction terms indicates that the higher confidence was not caused by lying at interview. If it were, we would expect the confidence for participants in the control condition to significantly differ from confidence in the lying conditions, since control participants did not lie at interview. This suggests that the increased confidence for commission errors is not because lying led subjects to falsely remember their lie as the truth.

The increased confidence for commission errors might be explained by the slight liberal response bias participants showed in all conditions. Responding incorrectly to a lie item meant that participants answered "yes" to these questions, whereas responding incorrectly to a truth item means that participants answered "no". Because participants showed a tendency to answer "yes" to all questions, they may have been more confident giving a "yes" response, regardless of whether it was the correct answer. Thus, a liberal response bias might have led to high confidence when they incorrectly stated that a lie was experienced in the mock crime.



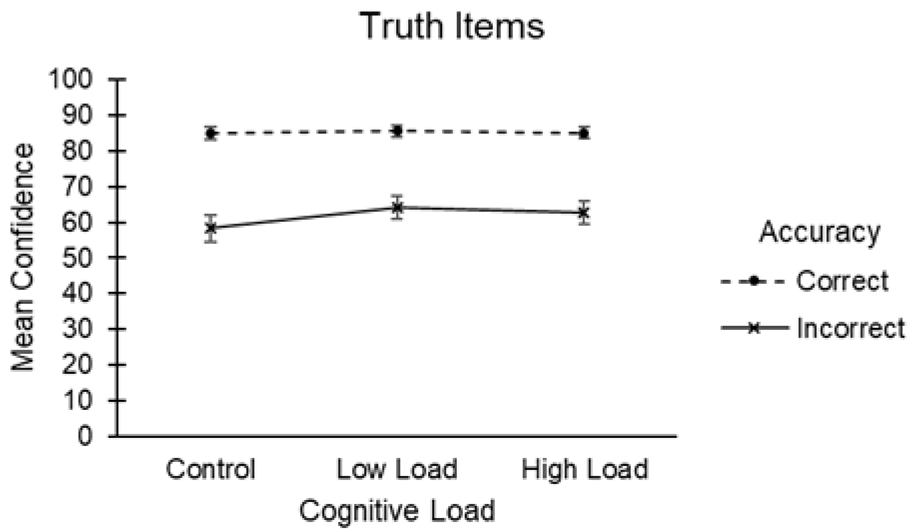


Figure 4.2. Mean confidence ratings (scale 0-100) for lie items (items subjects stated as a lie during interview [top panel]) and the corresponding truthful item from the mock crime [bottom panel]. Ratings are shown as a function of Cognitive Load and whether the item was correctly identified as a lie or the item from the mock crime in the recognition test (Accuracy). Error bars represent ±1 SE of the mean.

4.4.5 Discussion

Taken together, the findings of Experiment 4 suggest that lying at interview led people to remember fewer items from the mock crime (i.e., to make more omission errors) and to more often mistake their lies for the truth (i.e., to make more commission errors), regardless of the cognitive effort required to lie. However, the prediction that commission errors would be higher in participants who lied under low load was not supported. Commission errors were indeed marginally higher for participants who lied under low load, but this increase was not significant. The manipulation check suggests that the cognitive load manipulation was not strong enough to elicit significant differences between groups. Strengthening the cognitive load manipulation may therefore induce a significant difference in commission errors in the predicted direction.

The signal detection measures tell us that lying impaired participants' recognition ability, and therefore has real effects on people's memory, as opposed to their tendency to answer "yes" or "no" at test. Nonetheless, the cognitive effort of lying did not moderate the effect of lying on memory for any measures. Confidence was also unaffected by the cognitive effort of lying. Because the cognitive load manipulation was weak, the findings are inclusive regarding the implications for the MAD framework's cognitive load hypothesis.

Nonetheless, this experiment shows that lying by providing counterfactual information (fabrications) not only increased commission errors, but also omission errors. The fabrications in this experiment therefore led to a global memory impairment, not a selective increase in commission errors. Experiment 5 strengthens the cognitive load manipulation used in the present study to provide more conclusive results regarding the MAD framework.

4.5 Experiment 5

The aims of Experiment 5 were two-fold: First, to strengthen the cognitive load manipulation used in Experiment 4, and second, to investigate how manipulating different types of cognitive load might affect memory differently. As outlined in Section 1.2, there are many factors that contribute to the increased cognitive cost of lying occurring at each stage of deception, from the decision to lie to the lie's delivery. Regarding lie generation specifically, lying is typically more cognitively demanding to the extent that it requires manipulation of existing memories to construct a counterfactual answer (Sporer, 2016; Walczyk et al., 2014). Altering the difficulty of the anagram in Experiment 4 is analogous to increasing the extent to which people must manipulate available information to construct a fabrication. However, this is just one way that the cognitive effort of lying can be increased. Other methods of increasing cognitive load during deception may affect memory differently.

According to ADCAT (Walczyk et al., 2014), there are different types of cognitive load that might contribute to the additional demand of lying. The authors draw on *cognitive load theory* (Sweller, van Merrienboer, & Paas, 1998) to divide the cognitive load that liars experience into two categories: [1] *Intrinsic cognitive load*, defined as "the demand on cognitive resources inherent to deceive well" (Walczyk et al., 2014, p. 23). For instance, constructing elaborate lies is inherently more difficult than constructing simple lies. [2] *Extraneous cognitive load*, defined as "any situational factor external to the act of deception that reduces respondents' cognitive resources while lying" (Walczyk et al., 2014, p. 23). The most common way to increase extraneous load is to introduce a concurrent secondary task. For instance, asking interviewees to perform a driving simulation task occupies attentional and working memory processes that otherwise could be used to facilitate lying (Walczyk et al., 2014). The anagram method used in Experiment 4 therefore targeted intrinsic cognitive load by increasing the inherent difficulty of lie generation. But it may be that increasing liars' intrinsic load has different effects on memory compared to increasing extraneous load. Indeed, previous research suggests that this may be the case.

There are three main ways that increasing the cognitive demand of fabrication might affect later memory. First, it might affect the liar's memory for the truth (*truth item memory*). This is the main remit of the MAD framework. Second, it might affect the liar's memory for the specific lie they told (*lie item memory*), and thirdly, it might affect the liar's memory for having lied at all (*source memory*). Broader memory research suggests that increasing the cognitive demand of fabrication could affect all three types of memory. Moreover, the way in which memory is affected may depend on the type of cognitive load that is targeted – intrinsic or extraneous.

4.5.1 The potential effect of increasing liars' intrinsic cognitive load

Previous studies suggest that increasing the intrinsic cognitive load of a task promotes better encoding of source and improves later source attributions accordingly. This has primarily been studied in the context of the *generation effect* – the finding that memory performance is better for information that participants generate themselves compared to information they merely read (Bertsch et al., 2007; Slamecka & Graf, 1978). Such research shows that source memory (but not necessarily item memory) is more accurate for items that are hard to generate compared to items that are easy to generate (Nieznański, 2011). In other words, when the intrinsic cognitive demand of generation is higher, participants are better at source monitoring. Related research has shown that participants are better at identifying whether they saw or merely imagined a shape when imagining the shape was harder compared to when it was easier (Finke, Johnson, & Shyi, 1988). The idea is that the more effortful generation is, the more cognitive operations are recruited to complete the task, and so the more cognitive operations are available to aid source decisions at test (Dewhurst & Hitch, 1999; Hicks & Marsh, 1999).

Returning to deception, a fabrication will be intrinsically more cognitively demanding when it is harder to generate than when it is easy to generate. If a fabrication is difficult to concoct, it will not 'spring to mind', but rather will require the liar to search long term memory and manipulate existing details to form a counterfactual answer (Sporer, 2016). These processes – retrieval, manipulation, and organization – are the exact cognitive operations that the SMF posits will later help us to identify information in memory as self-generated rather than experienced (Johnson et al., 1993; McDonough & Gallo, 2008). A questioning procedure that forces liars to recruit more cognitive operations might therefore improve later source memory, meaning that the individual is more likely to remember that they lied. This in turn might make it easier for the individual to discriminate lies from truths later on. Commission errors could therefore decrease when the intrinsic cognitive load of fabrication is higher, contrary to the MAD framework's hypothesis. Moreover, because such cognitive operations are likely to be recruited less often when participants generate easier fabrications, source monitoring could be poorest—and commission errors highest—when people fabricate under low cognitive load. Thus, contrary to the predictions born out of the MAD framework, fabricating under low load may increase commission errors, whereas fabricating under high intrinsic load may decrease them (as predicted in Experiment 4).

4.5.2 The potential effect of increasing liars' extraneous cognitive load

Increasing the extraneous demand of fabrication might instead have negative effects on memory. Requiring participants to complete a secondary task divides attention and disrupts encoding so that a weaker memory trace is formed (Troyer & Craik, 2000). Consistent with this, numerous studies show that introducing a secondary task at encoding impairs later memory performance (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Lane, 2006; Pérez-Mata, Read, & Diges, 2002; Peters et al., 2008). This has been studied using both misinformation and the Deese-Roediger-McDermott (DRM) paradigms. In the DRM paradigm, participants learn lists of words that are semantically associated with an unpresented word. In later recall and recognition tests, participants often falsely remember the unpresented word, despite not having initially learnt it (Roediger & McDermott, 1995). When participants undertake a secondary task while learning the word lists, false recall rates increase further. For example, in one study, participants completed an articulatory suppression task, which required them to repeat the words “coca cola” out loud while learning the word lists. The authors proposed that the secondary task divided attention during encoding so that participants processed the words more superficially. Participants therefore failed to encode key item-specific information that would help them monitor the source of their memories at test. Thus, the addition of a secondary task impaired source monitoring (Van Damme, Menten, & d’Ydewalle, 2010).

Studies using the misinformation paradigm have shown similar effects. In one study, participants witnessed a mock crime under full attention or while completing a secondary music task. They then answered misleading questions about the event before their source memory was assessed. Participants were significantly more likely to incorporate the misleading post-event information into memory when they encoded the original event under divided attention compared to under full attention (Lane, 2006). This is thought to be because dividing attention prevented participants from encoding the visual, spatial and contextual details that would later assist their source monitoring decisions.

Returning to deception, the aforementioned research suggests that dividing attention while participants fabricate might disrupt their encoding of the interview and result in poorer memories for what they lied about (source memory) or the specific lies they told (lie item memory). This could have knock-on effects for the liar’s memory of the truth, but conflicting hypotheses can be formed about whether truth item memory will be impaired or improved. On the one hand, we know that poorer source memory for having lied increases the likelihood that participants will incorporate their lies into memory (Polage, 2004). Thus, if increasing extraneous cognitive load impairs source memory, commission errors may increase. On the other hand, divided attention while lying might prevent participants from visualizing their lies, so they generate fewer perceptual details that would lead them to misattribute the source of their memories later on. This might make it easier for them to distinguish their memories of the truth (encoded under full attention) from memories of their lies (encoded under divided attention). Thus, if increasing extraneous cognitive load impairs lie item memory, commission errors may decrease.

In sum, increasing the extraneous cognitive load of fabrication by requiring participants to complete a secondary task could affect all three types of memory: Source memory, lie item memory, and truth item memory. By disrupting encoding of the interview, the individual might later have a poorer memory for their lies and this in turn may affect how well they remember the truth. Given that this could increase or decrease commission errors, no specific hypothesis can be developed regarding support for the MAD framework.

Why is it important to understand how increasing different types of cognitive load during deception affects later memory? The main reason is that cognitive load is not only an important concept in the MAD framework, but also forms an entire approach to lie detection that is popular amongst deception researchers (Vrij et al., 2017). If manipulating interviewees' cognitive load can affect their memory, this could have implications for lie detection approaches that rely on increasing cognitive load.

4.5.3. Implications for lie detection

Several researchers have endorsed the idea that the cognitive burden of lying can exploited to assist lie detection (Vrij, Fisher, et al., 2008; Walczyk et al., 2012, 2014) and this has been termed the *cognitive approach to lie detection* (CALD; Vrij & Fisher et al., 2008). CALD recommends techniques that further increase the interviewee's cognitive load during questioning. Liars typically have fewer residual cognitive resources than truth tellers, so further increasing cognitive demand should lead to cognitive overload in liars, which magnifies otherwise faint and unreliable cues to deception, such as slower response times, fewer details, and increased inconsistencies (Vrij et al., 2011; Vrij, Mann, et al., 2008).

Consistent with *cognitive load theory (Sweller et al., 1998)*, lie detection can be facilitated by increasing the intrinsic or extraneous cognitive load of lying. Researchers have tested several ways to increase the intrinsic cognitive load of lying. One of the first attempts was *time restricted integrity-confirmation* (*TRI-Con*), a questioning procedure developed by Walczyk et al., (2005). TRI-Con attempts to make the truth more salient in an individual's memory and requires rapid responses to interrelated unanticipated questions that may induce contradictions. Asking unanticipated questions is an approach unto itself that prevents liars from rehearsing their stories, thus increasing the intrinsic demand of lying and, in turn, cues to deceit (Ioannou & Hammond, 2015; Shaw et al., 2013).

Extraneous cognitive load is typically increased by asking interviewees to perform a secondary unrelated task while they undergo questioning. Because working memory capacity is constrained, interviewees may experience cognitive overload—and impaired performance—if attempting additional tasks when working memory is already at capacity (Sweller et al., 1998). Indeed, research has shown that when people lie, they increase their driving speed while operating a driving simulator (Gawrylowicz et al., 2016) or sort fewer objects in a haptic sorting task (Lancaster, Vrij, Hope, & Waller, 2013).

Taken together, the previous sections suggest that increasing the cognitive demand of lying might affect memory differently depending on how cognitive load is increased. Overall, existing evidence suggests that increasing intrinsic cognitive load might benefit memory, whereas the anticipated effect of increasing extraneous cognitive load is less clear. This has potentially important implications for lie detection in interview settings.

An individual who does not remember the lies they have told or the fact that they lied may be less consistent if they undergo repeated questioning. Inconsistency is a cue to deception (Hartwig, Granhag, Strömwall, & Vrij, 2005; Leins, Fisher, Vrij, Leal, & Mann, 2011; Vredeveldt et al., 2014). Increasing the intrinsic demand of lying might therefore help the liar by making it easier for them to be consistent over time, whereas increasing extraneous cognitive load might hinder the liar by increasing the likelihood of inconsistencies over multiple interviews. It is therefore important to investigate the interaction between lying and memory, as this might influence the efficacy of the cognitive approach to lie detection in interview settings.

4.5.4 Method summary

To begin to understand the complex interactions between cognitive load, memory, and lie detection, Experiment 5 manipulated the intrinsic and extraneous cognitive load of lying to determine their influence on source and item memory. As in Experiment 4, participants were provided with fabrications to approximate a situation where an individual knowingly reports false details that they have heard elsewhere or been told to say (e.g., an intimidated witness). The experiment was conducted over two sessions. In Session 1, participants witnessed a mock crime before undergoing questioning. All participants lied about some items and told the truth about others under either low cognitive load, high intrinsic cognitive load or high extraneous cognitive load.

Once again, anagrams were used to manipulate the extent to which participants were involved in lie generation. Participants in the low load condition received the lie as an easy anagram, whereas participants in the high intrinsic load condition received the lie as a hard anagram. Participants therefore had to put more mental effort into creating the lie itself and should therefore have experienced greater intrinsic cognitive load than subjects in the low load condition. The anagram difficulty was further increased compared to those used in Experiment 5 to ensure that the manipulation worked.

Participants in the high extraneous load condition received the lie as an easy anagram, but completed a concurrent articulatory suppression task that required them to repeat aloud a string of nonsense syllables. This is a theoretically informed choice of secondary task recommended by Walczyk et al. (2013). According to Walczyk et al., articulatory suppression occupies the phonological loop, a subsystem of working memory that is considered critical for successful deception (Sporer, 2016). The phonological loop temporarily stores verbal information and keeps it active via vocal or subvocal rehearsal (Baddeley, 2012) and is taxed significantly more during lying than truth telling (Sporer, 2016). Thus, if the phonological loop is occupied using articulatory suppression, the executive supervisory system must prioritize engagement in either the articulatory suppression task or lying. Articulatory suppression therefore increases extraneous cognitive load by targeting the same cognitive resources required for lying.

In Session 2 of the experiment, participants completed a recognition and source test to assess their item and source memory for the details of the mock crime. The study was preregistered on the Open Science Framework at the following link: https://osf.io/ydt2m/?view_only=1cb24ce5a377441fa746628d242910ed

It was predicted that: [1] Participants in the high intrinsic load condition will make fewer commission errors and show better discriminability for mock crime items that they previously lied about than for items they did not speak about (control items). In other words, increasing the intrinsic cognitive load of lying will improve memory relative to not rehearsing the details. [2] Participants in the low load condition will make more commission errors and show poorer discriminability for mock crime items that they previously lied about than for control items. In other words, low cognitive effort when lying will impair memory relative to not rehearsing the details, and [3] Source memory performance will be highest for participants in the high intrinsic load condition and lowest for participants in the high extraneous load condition.

4.6 Method

4.6.1 Participants.

The experiment was completed by 180 adults in exchange for course credit or a small monetary reward (mean age = 18.72 years, *SD* = 1.62 years; 21 male, 158 female, 1 other). Participants were primarily native English speakers (n = 133). Of those whose first language was not English, 29 rated themselves as fluent to a native or excellent standard and the remaining participants rated their English as either good (n = 10), satisfactory (n = 4) or poor (n = 4).

To determine an appropriate sample size, a power analysis was conducted using G*Power 3.1.9.2 (Faul et al., 2007). Experiment 4 found a medium effect size for the difference in memory performance for items participants lied about versus control items. The power analysis was therefore based on a medium effect size of *Cohen's d* = 0.5 (*f* = 0.25), which revealed that the present experiment requires a total of 158 participants to achieve a power of 0.8 with an alpha level of .05. The study was approved by the psychology department ethics committee at the University of Warwick.

4.6.2 Stimuli and apparatus.

Mock crime. The experiment used a mock crime video developed for use in eyewitness research (Takarangi, Parker, & Garry, 2006). This soundless video depicts an electrician, Eric, snooping around somebody's house and stealing various items while on the job. Twenty critical details were selected from the video for participants to memorize (18 critical, 2 practice), for example, that Eric found the key under a *flowerpot,* stole *earrings,* and drank a can of *coke*.

To ensure that participants encoded the critical items well, the video was edited using PsychoPy (Peirce, 2007) to include snapshots of these items. After each item was featured in the video, a snapshot image appeared on screen, together with a label, for 5 seconds. Four attention checks were also included to assess whether participants paid attention to the video. Each check consisted of a multiple-choice question concerning a detail that immediately preceded it (details unrelated to target items). Participants selected one of 5 possible answers and were given feedback. If they answered incorrectly, they were told the correct answer and asked to pay more attention to the video. The attention checks were placed at fixed intervals in the video: The first at 50 seconds, the second at 2 minutes, the third at 3 minutes 45 seconds and the fourth at 7 minutes 20 seconds. Together with the snapshots and attention checks, the video's duration was 9m 6s.

Lie creation. As in Experiment 4, participants were provided with the lies to deliver during interview, therefore mimicking a situation where an individual knowingly reports false details that they have heard elsewhere or been told to say (e.g., an intimidated witness). A corresponding lie was therefore created for each of the 20 critical items from the mock crime, for example *doormat, necklace,* and *Pepsi* were the corresponding lies for the items *flowerpot*, *earring,* and *coke*.

As in Experiment 4, participants were presented with the answer to each interview question in anagram form. Each answer was presented as either an easy anagram that required participants to switch two underlined letters, or a hard anagram that required participants to rearrange all, or most, letters. The full stimulus set is shown in Appendix 7.

4.6.3 Design and procedure.

The experiment was a 3 (Cognitive Load: low load, high intrinsic load, high extraneous load) x 3 (Veracity: lie, truth, control) mixed design. Cognitive Load was the between-subjects variable. The experiment involved 2 sessions, completed 3 weeks apart. Session 1 consisted of a mock crime, distraction and interview phase and was completed in individual laboratory cubicles. Session 2 consisted of a recognition and source test and was completed online. To prevent participants from ascertaining the study's purpose, they were informed that the aim of the experiment was to examine people's problem-solving skills and memory when lying under pressure. A suspicion check was included to determine if any participants ascertained the study's true purpose.

Session 1

Mock crime phase. After providing consent and demographic information, participants viewed the mock crime video. Participants were asked to imagine that they had an electrician friend, Eric, who they accompany on a job. While on the job, they witness Eric taking certain liberties and stealing several items. Participants were instructed to imagine that they were with Eric witnessing the events unfolding and to memorize the items that appear as snapshots. They were also informed that there would be attention checks at random intervals.

Distractor phase. This was identical to Experiment 4.

Interview phase. Next, participants were informed that Eric had been arrested and that they were to be questioned about the activities they saw him perform. They were informed that they would be provided with the answers as anagrams that they must solve and deliver, and told that they will give a mixture of deceptive and truthful answers to protect Eric, but avoid arousing suspicion. The interview addressed 12 of the 18 critical items from the video and the question order was randomized across participants. The remaining 6 items were not addressed during interview and therefore served as control items to determine baseline memory. Each question was written on screen and participants were provided with the answers in anagram form to solve and deliver. Half of the answers were lies, highlighted in yellow, and half were truths, highlighted in grey. The answers were highlighted in their respective colors so that participants were sure which were truthful and which deceptive (as per Rindal, 2017). Participants were given 30 seconds to type their response. If a response was not registered in this time, the program moved onto the next question.

Participants were randomly allocated to the *low load*, *high intrinsic load*, or *high extraneous load* condition. They were given full instructions followed by two example questions and answers, and then two practice questions that addressed the practice items. Feedback was given for these questions. Participants in the low load condition were given each answer in the form of an easy anagram that required them to switch the position of two underlined letters to reveal the answer. For instance, in response to the question "What item of jewelry did Eric steal from the first bedroom?", participants solved either the truthful anagram "EARIRNGS" or the deceptive anagram "NECLKACE". Participants in the high intrinsic load condition were given hard anagrams to solve. Again, participants rearranged underlined letters to reveal the answers, but either all or most of the letters were underlined. For example, the hard anagrams for the above question were "KNELCEAC" (necklace) and "GERARINS" (earrings).

Participants in the high extraneous load condition were given the same easy anagrams as those in the low load condition, but engaged in a concurrent articulatory suppression task while answering each question. Specifically, participants repeated a nonsense syllable sequence (‘bah-bay-bee-boo’; as recommended by (Walczyk et al., 2013) in time to a 120bpm metronome before the question and answer appeared. Participants were instructed to repeat the syllable sequence in time with the beat while they solve the anagram and type their response. Participants’ verbal responses were recorded to ensure that they complied with the articulatory suppression task. A self-paced break was taken in-between questions to enable participants to rest from repeating the syllable sequence.

All participants were instructed to give their answers as part of a full sentence, rather than merely report the solved anagram and to picture the item in their mind’s eye. Thus, for the above example, participants would answer with something like, “Eric stole a necklace/earrings from the first bedroom.” This was to encourage participants to integrate the answer with the question.

The specific items that participants told the truth, lied about, or were not asked about were counterbalanced so that every item served in each Veracity condition equally across participants. To achieve this, the 18 critical items were randomly allocated to one of 3 sets of 6 items. Participants lied about one set, told the truth about another and were not asked about the third.

On completing the interview, participants rated how much mental effort they required to answer the questions on a 9-point Likert scale ranging from “very low mental effort” to “very high mental effort” (as in Experiment 4).

Session 2.

Twenty-one days after completing Session 1, participants were emailed a link to complete Session 2 of the study online. Memory was assessed for the 18 critical items witnessed in the mock crime using a recognition and source test. Participants were instructed to answer all questions truthfully and completed 4 practice questions with feedback before beginning.

The recognition test was an identical format to that used in Experiment 4; that is, there were two questions per critical item—one containing the truthful answer (e.g., “Did Eric steal earrings from the first bedroom?”) and one containing the corresponding deceptive answer (e.g., “Did Eric steal a necklace from the first bedroom?”). For items that participants did not lie about, the question containing the deceptive answer functioned as a lure. Participants indicated “yes” or “no” to each question and rated their confidence on a scale of 0 = *completely uncertain* to 100 = *completely certain*.

To assess source memory, on the next page, participants were asked “Please choose an option that best describes your memory for the item ___” (whichever item was addressed on the previous page) and selected one of five options: 1) This was shown in the mock crime and I gave it as a truthful answer to a question at interview; 2) This was shown in the mock crime, but I lied when I was asked about it at interview; 3) This was shown in the mock crime, but I was NOT asked a question about it at interview; 4) This was NOT shown in the mock crime, but I gave it as a lie in response to a question at interview; 5) This was NOT shown in the mock crime and I was NOT asked a question about it at interview.

The source question for each item always followed the recognition and confidence questions for that same item, but the order in which items were addressed was randomized across participants. On completion of the memory test, participants were asked to state what they believed the study’s hypothesis was. No participants correctly guessed the study’s hypotheses, but 20 participants discerned that the experiment concerned some relationship between lying and memory.

4.7 Results

4.7.1 Preliminary analyses

Before addressing the primary research question of the experiment, four manipulation and performance checks were performed.

Engagement in mock crime. Participants’ performance in the attention checks indicates that they engaged appropriately with the mock crime. Attention check data from 28 participants were lost due to an equipment error. The remaining 152 participants passed an average of 88.82% of checks (*SD* = 16.22%). Only 1 participant failed more than 50% of attention checks and was excluded from all further analyses (as set out in the study pre-registration).[6]

Manipulation check. Participants' ratings of the mental effort required to complete the interview indicated that the cognitive load manipulation was successful. Mental effort ratings significantly differed across conditions ($F(2,175) = 55.71$, $p < .001$). Compared to those interviewed under low load ($M = 3.43$, $SD = 1.84$ [scale of 1-9]), participants reported using significantly more cognitive effort when interviewed under high intrinsic load ($M = 5.93$, $SD = 1.82$, $p < .001$, 95% CI[1.72, 3.29]) and high extraneous load ($M = 6.76$, $SD = 1.68$, $p < .001$, 95% CI[2.54, 4.13]), as indicated by Bonferroni-adjusted pairwise comparisons. Additionally, participants in the high extraneous load condition rated the task as significantly more effortful than did participants in the high intrinsic load condition ($p = .04$, 95% CI[0.04, 1.61]).

Interview performance. Participants' performance levels during the interview were good across all conditions. Participants in the low load and high extraneous load conditions achieved the highest performance, correctly solving an average of 97.70% ($SD = 15.0\%$) and 95.69% ($SD = 20.31\%$) of anagrams respectively. Participants in the high intrinsic load condition performed at a lower level, correctly solving an average of 79.78% ($SD = 40.19\%$) of anagrams. For incorrect answers, the corresponding memory data for that item were excluded from subsequent analyses. All participants in the high extraneous load condition complied with the articulatory suppression task by repeating the syllable sequence aloud for the duration of the interview.

Attrition across sessions. The average time that elapsed between completion of Sessions 1 and 2 was 22.39 days ($SD = 1.70$ days). The time between sessions did not significantly differ across conditions, as indicated by a one-way between-subjects ANOVA ($F(2,162) = .03$, $p = .97$). Data were excluded from 7 participants who failed to

[6] Because only 1/152 participants failed more than 50% of attention checks, the 28 participants for whom we did not have attention check data were included in subsequent analyses.

complete part 2, and a further 7 participants who failed to complete part 2 within the 4-week deadline specified in the pre-registration. Subsequent analyses are therefore based on data from 165 participants, which still exceeded the sample size target of 158 participants.

The primary research question is now addressed in three parts: Determining the extent to which increasing cognitive load while lying influenced participants' [1] recognition performance, i.e., their ability to accurately recognize details from the crime event, [2] confidence in their memory performance, and [3] source memory, i.e., their ability to remember the origin of each item and how they spoke about them during interview.

4.7.2 Recognition performance

As predicted, the type of cognitive load imposed during interview affected participants' recognition performance. Signal detection measures were used to assess recognition performance, as in Experiment 4. To recap, signal detection theory defines each participant's recognition performance with a HR and FAR. The HR is defined as the probability of correctly responding "yes" in truth trials (recognition questions containing an item from the mock crime, e.g., "Did Eric steal earrings?") and the FAR is defined as incorrectly responding "yes" in lie trials (recognition questions containing a lie. E.g., "Did Eric steal a necklace?").

As in Experiment 4, A' and B" were also calculated as a measure of discriminability and response bias respectively. Table 4.2 shows the descriptive statistics for each condition. To assess recognition performance, multilevel models were fitted to the data from each signal detection measure with Veracity and Cognitive Load as predictors. For each measure, I first describe the fit of the model to the data before interpreting the results.

FARs (commission errors): As a direct measure of commission errors, FARs are the primary measure for assessing the experimental hypotheses. Increased FARs represent increased commission errors.

Including Veracity as a predictor significantly improved the model fit, compared to an intercept-only model ($\chi^2(6) = 42.16$, $p < .001$), but adding the main effect of Cognitive

Load did not further improve the model fit ($\chi^2(8) = 3.98$, $p = .14$). The model was significantly improved when including the interaction term between Veracity and Cognitive load ($\chi^2(12) = 18.13$, $p = .001$), indicating that the effect of lying on commission errors depended on the cognitive load imposed during interview. Separate multilevel models were fitted to each cognitive load condition to break down this interaction.

For participants interviewed under low load, the model including Veracity was a significant improvement to an intercept-only model ($\chi^2(6) = 32.69$, $p < .001$). Furthermore, the model contrasts indicate that the prediction that commission errors (i.e., FARs) would increase was supported. As Table 1 shows, commission errors almost doubled for items that participants lied about under low load compared to control items ($b = .19$, $t(110) = 5.43$, $p < .001$), suggesting that lying under low cognitive load impaired memory for the truth. Commission errors did not increase for items that participants reported truthfully during interview compared to control items ($b = .01$, $t(110) = 0.29$, $p = .77$), indicating that the memory impairment was specific to lied-about items.

For participants interviewed under high intrinsic load, including Veracity was a significant improvement to an intercept-only model ($\chi^2(6) = 21.41$, $p < .001$). It was predicted that commission errors would decrease for participants who lied under high intrinsic load, but contrary to this prediction, commission errors actually increased by 50% (see Table 1). The model contrasts confirmed that this increase was significant $b = .12$, $t(110) = 3.20$, $p = .002$). Commission errors were unaffected for items participants told the truth about at interview ($b = -.06$, $t(110) = -1.51$, $p = .13$), indicating that this memory impairment was specific to lied-about items.

For participants interviewed under high extraneous load, it was reasoned that memory could be impaired or enhanced for lied-about compared to control items. Indeed, there was no effect on memory for participants who lied under high extraneous load: Including Veracity did not improve the fit of the multilevel model, relative to an intercept-only model ($\chi^2(6) = 0.81$, $p = .67$). Thus, lying under high extraneous load neither impaired or enhanced memory for lied-about items.

Although commission errors were highest for participants who lied under low load, the overall model including Veracity and Cognitive Load indicated that the increase in commission errors relative to control items was not significantly greater for participants interviewed under low load ($M = .19$, $SD = 0.27$) compared to participants interviewed under high intrinsic load ($M = .11$, $SD = .27$; $b = -.07$, $t(322) = -1.34$, $p = .18$). Lying under low load did not therefore lead to a greater memory impairment than lying under high intrinsic load.

HRs (omission errors). HRs can be analyzed to assess omission errors—incorrect "no" responses in truth trials represent omission errors and a lower HR therefore indicates increased omission errors.

Including Veracity as a predictor significantly improved the model fit, compared to an intercept-only model ($\chi^2(6) = 71.53$, p < .001). Neither the main effect of Cognitive Load ($\chi^2(8) = 4.89$, p = .09) nor the interaction between Veracity and Cognitive Load further improved the model fit ($\chi^2(12) = 5.56$, p = .23). The model contrasts (containing Veracity as the single predictor), indicated that HRs were higher for items that participants truthfully reported during interview compared to control items ($M_{\text{Control}} = .69$, $SD = .22$; $M_{\text{Truth}} = .86$, $SD = .22$; $b = .17$, $t(326) = 8.32$, $p < .001$). Lying about an item during interview did not decrease HRs compared to control items ($M_{\text{Lie}} = .72$, $SD = .15$; $b = .03$, $t(326) = 1.37$, $p = .17$). Lying at interview did not therefore increase omission errors.

A' (discriminability). Analysis of the A' values determines whether the increase in commission errors when participants lied under low load or high intrinsic load can be attributed to impaired recognition memory, rather than differences in response bias. A' values were calculated in the same way as Experiment 4 (including the loglinear adjustment procedure for participants who achieved perfect performance (HR=1, FAR=0); see Section 4.4.3).

To recap, A' typically ranges from 0.5 (indicating chance performance) to 1 (indicating perfect performance). Table 4.2 shows that discriminability was well above chance for all conditions. Including Veracity as a predictor in the multilevel model significantly improved the model fit ($\chi^2(6) = 65.04$, $p < .001$), but the main effect of Cognitive Load did not ($\chi^2(8) = 3.35$, $p = .19$). However, adding the interaction term between Veracity and Cognitive Load significantly improved the model ($\chi^2(12) = 18.63$, $p < .001$), indicating that the effect of lying on discriminability was affected by the cognitive load imposed during interview. Once again, separate multilevel models were fitted to the data from each cognitive load condition to break down this interaction.

For participants interviewed under low cognitive load, the model was significantly improved when including Veracity compared to an intercept-only model ($\chi^2(6) = 29.38$, $p < .001$). As predicted, participants interviewed under low load showed impaired discriminability (i.e., lower A' values) for items they lied about compared to control items, as confirmed by the model contrasts ($b = -.09$, $t(110) = -3.09$, $p = .003$). Unsurprisingly, discriminability was improved for items participants reported truthfully during interview compared to control items ($b = .07$, $t(110) = 2.65$, $p = .009$).

For participants interviewed under high intrinsic load, the model including Veracity was a significant improvement to an intercept-only model ($\chi^2(6) = 41.18$, $p < .001$). Contrary to the prediction, but consistent with the FAR analysis, the model contrasts indicated that discriminability was significantly poorer for items participants lied about under high intrinsic load compared to control items ($b = -.08$, $t(110) = 3.05$, $p = .003$). Again, discriminability was improved for items participants reported truthfully during interview compared to control items ($b = .11$, $t(110) = 3.96$, $p < .001$). Thus, the increase in commission errors when participants lied under low load and high intrinsic load can be attributed to impaired recognition memory.

For participants interviewed under high extraneous load, the model including Veracity was a significant improvement to an intercept-only model ($\chi^2(6) = 9.49$, $p = .009$). The model contrasts indicate that discriminability did not significantly differ for items participants lied about compared to control items ($b = .04$, $t(102) = 1.35$, $p = .18$), but discriminability was improved for items participants told the truth about compared to control items (*b* = .08, *t*(102) = 3.11, *p* = .002).

In sum, while all participants experienced better memory for items they reported truthfully during interview, only those interviewed under high extraneous load did not experience impaired memory for items they lied about during interview.

B" (response bias). It is possible that the increased commission errors for participants interviewed under low load and high intrinsic load are not only due to impaired memory, but also to differences in response bias across groups; that is, participants' tendency towards answering "yes" or "no" regardless of their memory for that item. To determine if response bias contributed to the increased number of commission errors, participants' B" values were compared across conditions.

Table 4.2 shows that B" was close to 0 in all conditions, indicating very little response bias. This suggests that the differences in commission errors between groups are indeed attributable to changes in discriminability, rather than response bias. Nonetheless, response bias did significantly differ across conditions. Including Veracity as a predictor significantly improved the fit of the multilevel model compared to an intercept-only model ($\chi^2(6) = 16.87$, $p < .001$) and including Cognitive Load further improved the model fit ($\chi^2(8) = 6.01$, $p = .05$). Including the interaction term between Veracity and Cognitive Load did not further improve the fit ($\chi^2(12) = 1.71$, $p = .79$).

For the main effect of Veracity, participants showed a slightly liberal response bias in favor of stating that items they were interviewed about were shown in the mock crime ($M_{lie\ items} = -0.08$, $SD = 0.25$; $M_{truth\ items} = -0.05$, $SD = 0.31$) and a slightly conservative bias in favor of stating that items they were not interviewed about (control items) were not shown in the mock crime ($M = 0.04$, $SD = 0.25$). The model contrasts confirmed that participants were significantly more liberal for both lied-about items compared to control items (*b* = -0.12, *t*(326) = -4.02, $p < .001$) and truth items compared to control items *b* = -0.09, *t*(326) = -2.83, *p* = .005). Put simply, questioning participants about an item meant they were more likely to state that the item was shown in the crime, regardless of whether they remember so.

For the main effect of Cognitive Load, participants interviewed under high extraneous load were significantly more liberal than participants interviewed under low load ($M_{high-ext}$ = -0.08, *SD* = 0.28; M_{low} = -0.009, *SD* = 0.27; *b* = -0.07, *t*(162) = -2.15, *p* = .03). A liberal bias increases participants' tendency to respond "yes" and a strong enough liberal bias would therefore increase commission errors. Interestingly, participants interviewed under high extraneous load were the only group that did not see an increase in commission errors after lying. Thus, despite showing the strongest liberal response bias, these participants still made the fewest errors after lying at interview. Response bias did not significantly differ for participants interviewed under high intrinsic load compared to participants interviewed under low load ($M_{high-int}$ = -0.01, *SD* = 0.28; *b* = -0.001, *t*(162) = -0.02, *p* = .98).

Table 4.2

Mean hit rate (HR), false alarm rate (FAR), A prime (A') and response bias (B") for each type of item addressed during interview as a function of cognitive load.

	HR			FAR			A'			B"		
	C	L	T	C	L	T	C	L	T	C	L	T
LL	0.67 (0.23)	0.68 (0.22)	0.83 (0.17)	0.22 (0.19)	0.41 (0.24)	0.23 (0.21)	0.75 (0.17)	0.67 (0.22)	0.83 (0.11)	0.08 (0.27)	-0.09 (0.23)	-0.02 (0.27)
HLI	0.70 (0.20)	0.71 (0.23)	0.89 (0.23)	0.22 (0.20)	0.34 (0.24)	0.18 (0.19)	0.77 (0.16)	0.69 (0.19)	0.87 (0.09)	0.06 (0.29)	-0.06 (0.18)	-0.03 (0.32)
HLE	0.70 (0.22)	0.78 (0.20)	0.86 (0.15)	0.30 (0.18)	0.30 (0.20)	0.28 (0.21)	0.74 (0.15)	0.78 (0.16)	0.83 (0.13)	-0.03 (0.27)	-0.11 (0.33)	-0.10 (0.31)

Note: Numbers in parentheses are standard deviations. LL = Low load, HLI = High Intrinsic Load, HLE = High Extraneous Load. 'C' represents items participants were not asked about during interview, 'L' represents items participants lied about during interview, and 'T' represents items participants reported truthfully during interview.

4.7.3 Confidence

How confident were participants when they mistakenly reported a lie item as originating from the mock crime (i.e., committed a commission error)? Increased confidence in such memories may be interpreted as evidence that participants really did incorporate their lies into memory.

Confidence ratings for the mock crime items that participants lied about, the lies reported, and mock crime items that participants reported truthfully were compared to confidence ratings for control items. Additionally, as in Experiment 4, a new factor (Accuracy*)* was created that split the confidence ratings into two levels: Ratings for recognition questions answered correctly and ratings for recognition questions answered incorrectly. Thus, the final multilevel model included three factors: Accuracy (correct, incorrect), Veracity (control items, lies reported, items lied about, items truthfully reported) and Cognitive Load (low load, high intrinsic load, high extraneous load). Analysis of confidence ratings was exploratory; there were no specific predictions about how lying under different levels of cognitive load would affect confidence in memory accuracy.

Model fitting. Including the main effect of Accuracy significantly improved the model fit compared to an intercept-only model ($\chi^2(6) = 16.87$, $p < .001$) and including the main effect of Veracity further improved the model fit ($\chi^2(9) = 25.04$, $p < .001$). However, the main effect of Cognitive Load did not further improve the model fit ($\chi^2(11) = 0.40$, $p = .82$). The model was significantly improved including the interaction term between Accuracy and Veracity ($\chi^2(14) = 91.12$, $p < .001$), but not from the further addition of any other interaction terms. Cognitive Load was therefore removed from the model, since neither the main effect of Cognitive Load, nor any of its interactions, significantly improved the model fit. The final model therefore included the main effects of Accuracy, Veracity and the interaction term between Accuracy and Veracity.

Model interpretation. Cognitive Load did not improve the model fit, indicating that, unlike recognition, confidence was unaffected by the cognitive load imposed during interview. Regardless of cognitive load, we can see from Figure 4.3 that participants were generally more confident in their memories when they were correct compared to incorrect, except for when judging their confidence for the lies they reported during interview. This is represented by the significant interaction between Accuracy and Veracity. To follow up this interaction, the data were split according to Accuracy and separate multilevel models were fitted to the data for correct and incorrect answers respectively.

The primary interest here is participants' confidence when they made commission or omission errors, and therefore when they responded incorrectly. For incorrect responses, including Veracity as a predictor significantly improved the model fit compared to an intercept-only model ($\chi^2(7) = 37.16$, $p < .001$). Participants' confidence when they made commission errors is indicated by responding incorrectly to questions containing a lie reported during interview (this indicates that they selected “yes”—the item was shown in the mock crime). The model contrasts showed that when participants made commission errors, they were significantly more confident in their memories ($M = 68.97$, $SD = 19.37$) than for their memories of control items ($M = 55.08$, $SD = 21.23$; $b = 13.71$, $t(329) = 5.67$, $p < .001$). Thus, despite mistaking lies for the truth, participants were more confident in the accuracy of their memory, suggesting that they may have believed that some of their lies were in fact the truth.

Participants' confidence when they made omission errors is indicated by responding incorrectly to questions containing items they lied about during interview (this indicates that they selected “no”—the item was *not* shown in the mock crime). The model contrasts showed that when participants made omission errors, they were significantly more confident in their memories ($M = 61.69$, $SD = 23.39$) than for their memories of control items ($b = 6.09$, $t(329) = 2.46$, $p = .01$). This suggests that lying weakened participants' memory for the truth.

For correct responses, including Veracity as a predictor significantly improved the model fit ($\chi^2(7) = 116.49$, $p < .001$). The model contrasts indicated that when participants correctly stated that the lies reported were not shown in the mock crime, their confidence was significantly lower ($M = 67.41$, $SD = 19.13$) than for control items

(M = 72.10, SD = 17.79; b = -4.71, $t(486) = -3.87$, $p < .001$). Taken together, these findings suggest that lying made participants more certain that their memories were correct when in fact they were not, and less certain of the truth even when they correctly remembered it.

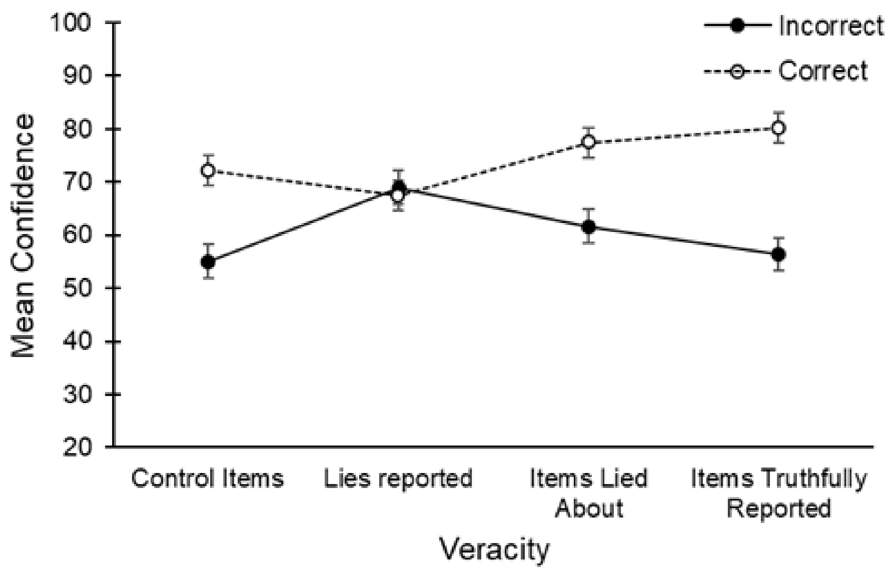


Figure 4.3. Mean confidence on a scale of 0-100 as a function of accuracy of recognition memory (responded correctly vs incorrectly) and Veracity. Error bars represent ±1 SE of the mean.

4.7.4 Source memory

By analyzing the types of source errors that participants made, we can assess how manipulating cognitive load during interview not only affects memory for the mock crime itself, but also memory for the interview. If manipulating cognitive load affects an individual's ability to later remember the lies told during interview, they might be less consistent when undergoing repeated questioning. Since inconsistency is often used as a cue to deception (Vredeveldt et al., 2014), an individual who forgets that they lied or the lies they told may be easier to detect in the course of an investigation. Accordingly, only the source memory for the items that participants lied about during interview are reported here (the detailed analyses for all item types is detailed in Appendix 9).

To assess source memory, a Poisson loglinear generalized estimating equation (GEE) analysis was conducted to examine the distribution of source responses for the lies reported during interview. From this, the likelihood that participants selected the correct source option compared to the remaining four source options was calculated to identify the types of source errors participants made in each cognitive load condition.

Figure 4.4 shows the distribution of source attributions. The correct source option for lies reported during interview is 4, which represents "this item was not shown in the mock crime, but I gave it as a lie during interview." Source attributions of 1, 2 or 3 represent the judgement that the item was shown in the mock crime, whereas attributions of 4 or 5 represent the judgement that the item was not shown in the mock crime. Since these items were not shown in the mock crime, a source attribution of 3 or below suggests that participants may falsely remember the item as originating from the crime.

We can see clear differences in the number of times that participants selected each source option for each condition. The GEE analysis revealed a significant interaction between Cognitive Load and the source option selected ($\chi^2(8) = 20.06$, $p = .01$). To break down this interaction, separate GEE models for each cognitive load condition were evaluated. For participants interviewed under low load, we can see from Figure 4.4 that the distribution of source responses is relatively evenly spread across the five options, with participants often selecting an option of 3 or below. This suggests that participants interviewed under low load experienced source confusion, as predicted. Indeed, all other source options were no less likely to be selected than the correct source option (all *p*s > .05; see Table 4.3 for odds ratios). Participants interviewed under low load therefore found it difficult to remember what items they had falsely reported during interview and often misremembered these items as coming from the mock crime.

Conversely, participants interviewed under high load did not show source confusion to the same extent for the lies they reported during interview. Participants interviewed under high intrinsic load were the only group to select the correct option significantly more often than all other options (see Table 4.3 for odds ratios). Thus, as predicted, these participants showed the best source memory performance.

Interestingly, participants interviewed under high extraneous load selected option 5 (“This item was not shown in the mock crime and I was not asked a question about it at interview”) for lie items most often, suggesting that they correctly remembered that the item was not shown in the crime, but did not remember reporting the lie at interview. Indeed, there was no significant difference in the odds that participants selected this option compared to the correct option (see Table 4.3). Thus, dividing attention led participants to often forget the lies they had told during interview.

In sum, interviewing participants under low load led to source confusion, where participants often misattributed lie items to the mock crime and failed to remember how they spoke about these items during interview. Interviewing participants under high extraneous load did not increase misattributions of lie items to the crime, but did lead participants to forget the lies reported during interview. Source memory was most accurate for participants interviewed under high intrinsic load.

Table 4.3

Odds ratios obtained from GEE analyses indicating the change in count for each source option compared to the correct source option (4) for the lies reported during interview.

Source Option	LL	HLI	HLE
1	0.73 [0.45, 1.20]	0.35** [0.21, 0.59]	0.54* [0.35, 0.83]
2	1.24 [0.82, 1.89]	0.53** [0.37, 0.76]	0.73 [0.49, 1.09]
3	0.79 [0.49, 1.27]	0.56* [0.37, 0.85]	0.50* [0.33, 0.76]
4	-	-	-
5	1.19 [0.80, 1.78]	0.61* [0.39, 0.94]	1.18 [0.79, 1.74]

Note: * indicates $p < .05$; ** indicates $p < .001$. ORs of <1 indicate that the option was selected less often than the correct option; ORs of >1 indicate that the option was selected more often than the correct option; ORs of <1 indicate that the option was selected less often than the correct option. Values in parentheses are 95% confidence intervals. LL = low load, HLI = high intrinsic load, HLE = high extraneous load. The source option numbers correspond to the following statements: 1 = “This was shown in the mock crime and I gave it as a truthful answer to a question at interview”, 2 = “This was shown in the mock crime, but I lied when I was asked about it at interview”, 3 = “This was shown in the mock crime, but I was NOT asked a question about it at interview”, 4 = “This was NOT shown in the mock crime, but I gave it as a lie in response to a question at interview”, 5 = “This was NOT shown in the mock crime and I was NOT asked a question about it at interview.”

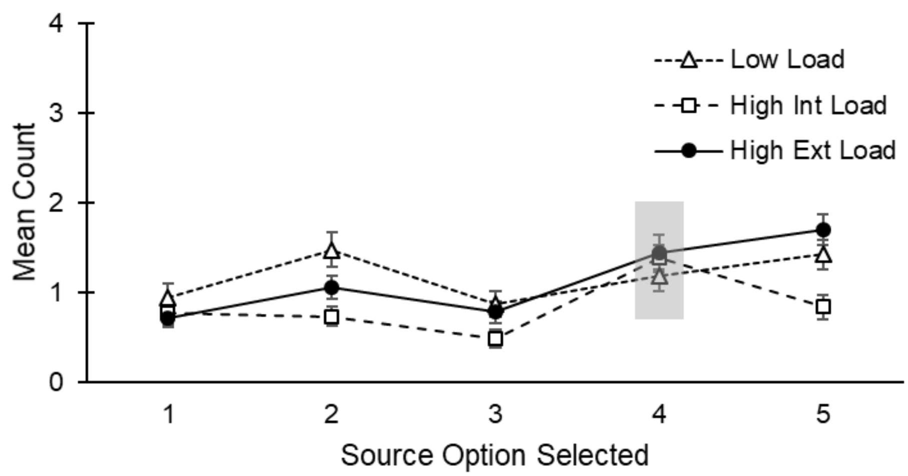


Figure 4.4. The distribution of source attributions for the lies that participants reported during interview. The correct source option is highlighted in grey. Attributions of ≤3 represent the judgement that the item was shown in the mock crime and attributions of >3 represent the judgement that the item was not shown in the mock crime: 1 = “This was shown in the mock crime and I gave it as a truthful answer to a question at interview”, 2 = “This was shown in the mock crime, but I lied when I was asked about it at interview”, 3 = “This was shown in the mock crime, but I was NOT asked a question about it at interview”, 4 = “This was NOT shown in the mock crime, but I gave it as a lie in response to a question at interview”, 5 = “This was NOT shown in the mock crime and I was NOT asked a question about it at interview.” Error bars represent ±1 SE of the mean.

4.8 Results summary

Based on cognitive load theory and research into the factors affecting source monitoring ability, it was reasoned that imposing different types of cognitive load might affect memory differently. Experiment 5 therefore investigated how lying under low cognitive load, high intrinsic load, or high extraneous load affects recognition memory, confidence and source monitoring for the details of a mock crime.

As predicted, the effect of increasing liars' cognitive load on memory differed depending on how cognitive load was manipulated. Lying under low load increased commission errors, led to source confusion, and impaired participants' ability to discriminate lies from truths in memory. Participants often mistook items they lied about for items that were shown in the mock crime, suggesting that lying under low load can promote false memories.

The prediction that increasing liars' intrinsic load would enhance source monitoring and decrease commission errors was partially supported. Interestingly, there was a dissociation between recognition and source memory when participants lied under high intrinsic load: These participants showed the best source performance, as expected, but commission errors increased, rather than decreased. This dissociation is discussed in the next section.

It was reasoned that memory performance for participants who lied under high extraneous load could go either way. In fact, these participants showed no change in commission errors or discriminability for items they lied about compared to control items, suggesting that dividing attention during interview protected participants from incorporating their lies into memory. The source memory data may explain why. As predicted, participants interviewed under high extraneous load showed the poorest source memory performance. Specifically, these participants often failed to remember the lies reported during interview. Thus, dividing attention might have prevented participants from adequately encoding their lies to the extent that there were fewer lies in memory to later confuse with the truth. The implications of these findings for the MAD framework and CALD are discussed in the following section.

4.9 General discussion

There were 3 main aims of Experiments 4 and 5: [1] To test the challenge to the MAD framework that the type of lie told is a distinct variable from the cognitive effort of lying, and therefore provide evidence that these variables should not be considered interchangeable predictors of the memory errors that might follow deception; [2] To test the hypothesis that increasing the cognitive effort of fabrication will decrease commission errors; and [3] To test the hypothesis that different methods of increasing the cognitive effort of fabrication will affect memory differently. Overall, the findings of Experiments 4 and 5 support aims [1] and [3], but not [2].

Experiment 4 showed that participants who knowingly reported fabricated answers in response to interview questions concerning a mock crime subsequently made more commission and omission errors than participants who did not undergo questioning. However, increasing participants' cognitive load during interview by increasing the difficulty of lie generation did not affect the error rate. Providing fabricated responses at interview therefore led participants to more often mistake those fabrications for the truth, as well as forget the truth, regardless of the cognitive effort required for lie generation. However, the manipulation check suggested that the cognitive load manipulation was not strong enough to elicit differences across groups. Accordingly, while the pattern of results was consistent with the prediction that commission errors would be highest for participants who lied under low load, the rate of commission errors did not significantly differ across cognitive load conditions.

A stronger manipulation in Experiment 5 yielded a significant increase in commission errors when participants lied under low load. Furthermore, commission errors increased when participants interviewed under high intrinsic load, but not when they lied under high extraneous load. Thus, the cognitive effort of lying can be manipulated independently of the type of lie to produce different effects on memory, as proposed in Chapter 3. This supports the proposition that the cognitive effort of lying and the type of lie should not be considered interchangeable predictors of subsequent memory errors.

Neither Experiment 4 nor 5 supported the prediction that increasing the difficulty of lie generation would decrease commission errors. Based on the SMF, it was reasoned that increasing participants' intrinsic cognitive load during interview would increase the number of cognitive operations associated with the memory for the lie, thereby enhancing later source monitoring and decreasing commission errors accordingly. Interestingly, while participants interviewed under high intrinsic load did indeed show the best source monitoring performance, this did not translate to decreased commission errors. Similar findings have been shown in eyewitness misinformation and forced fabrication studies, where participants show more memory errors when tested with a yes/no recognition test compared to when they are directed to consider the source of their memories (Ackil & Zaragoza, 2011; Lindsay & Johnson, 1989).

The dissociation between item and source memory performance may reflect two different memory processes targeted by the recognition and source questions. Memory judgements can be based on two different processes: *Familiarity*—quick judgements determined by the extent to which an item in memory is activated—or *recollection*—slower and more considered judgements determined by a more stringent retrieval search process (Yonelinas, 2002). Questions that measure item-context associations, like the source test in Experiment 5, likely tap into recollection processes, whereas questions that measure item recognition, like the recognition test used in both experiments, reflect both recollection and familiarity (Yonelinas, 2002). It may be that forced-choice yes/no recognition questions prompted familiarity judgements and that the increased familiarity with the item from lying at interview led participants to select the "yes" response more often, thus increasing commission errors. Conversely, the more detailed source questions may have prompted greater consideration and therefore recollective judgements that led participants to more accurately reflect on their memories.

If the above interpretation is correct, this opens the possibility that different questioning methods during an investigative interview might affect the quality of the testimony obtained. For instance, directing interviewees to consider the source of their memories might encourage them to engage more accurate recollective processes, rather than make familiarity-based judgements, therefore mitigating deception-induced memory errors.

The findings of Experiment 5 help to develop our theoretical understanding of how lying can affect memory for the truth. To recap, the MAD framework categorizes different types of lie based on their cognitive effort and states that cognitive effort can help to predict the memory errors that might follow: The framework states that fabrication requires relatively more cognitive resources and is associated with increased commission errors, whereas false denials and feigned amnesia require relatively few cognitive resources and are not associated with increased commission errors (Otgaar & Baker, 2018). The MAD framework therefore treats the type of lie and cognitive demand of lying as interchangeable predictors of the types of subsequent memory errors, however the findings of Experiment 5 show that these are indeed independent variables that can interact to affect memory in different ways. Participants told the same type of lie (fabrications) but varying the cognitive load required to generate these fabrications led to different effects on memory: Fabricating under low load increased commission errors, while fabricating under high extraneous load did not. Furthermore, these findings show that the nature of the relationship between the cognitive effort of lying and commission errors depends on how cognitive load is manipulated. Thus, increasing the cognitive effort of lying does not necessarily increase commission errors, as proposed in Chapter 3.

The findings of both experiments suggest that the main determinant of commission errors is not the cognitive demand of lying, but simply the presence of alternative information in memory that can be confused with the truth at test. When participants lied under high extraneous load, the source analysis indicated that they encoded lies from the interview less often. Thus, there were simply fewer opportunities for participants to confuse the lie with the truth and commission errors did not increase accordingly. Conversely, when participants lied under low load or high intrinsic load, they more effectively encoded the lies from the interview, increasing the chances that a lie can be confused with the truth, and commission errors increased. Thus, regardless of cognitive load, commission errors increased only when participants had to discriminate between a lie and a truth in memory. Increasing extraneous load appears to have protected participants only by reducing encoding of the lies told at interview and therefore eliminating this discrimination task.

Going forwards, the MAD framework may benefit from explicitly considering the cognitive effort of lying as a separate variable that can moderate the effects of different types of lie on memory. This allows the generation of novel hypotheses for future research, for instance by considering how manipulating the cognitive effort of false denials may affect the rate of omission errors.

The finding that increasing different types of cognitive load affected memory differently may have implications for lie detection techniques that encourage the imposition of cognitive load to magnify cues to deception, such as CALD. The findings of Experiment 5 give preliminary evidence that some methods of increasing cognitive load may be preferable to others. Specifically, methods that divide attention during interview may best preserve memory for the original event, so that an accurate account is retrievable if the individual later retracts their dishonest account. In addition, the findings suggest that dividing attention could facilitate lie detection by impairing encoding of the interview: Participants who lied under high extraneous load often forgot the lies they told at interview, which could promote inconsistencies in repeated interviews throughout an investigation, thus exposing deception (Vredeveldt et al., 2014). Conversely, questioning procedures that interfere with the inherent task of lying and encourage the generation of false information might simply create more opportunities for source confusion, and in turn, reduce the likelihood that a revised statement is accurate.

Future research should examine if the same pattern of results is found when using the specific procedures advocated by CALD, for instance increasing extraneous load by forcing participants to maintain eye contact with the interviewer (Vrij, Mann, Leal, & Fisher, 2010) and increasing intrinsic load by requesting that the interviewee provide their account in reverse chronological order (Vrij, Mann, et al., 2008). If the findings of Experiment 5 generalize to these techniques more broadly, some techniques of imposing cognitive load on interviewees may have detrimental effects on memory and therefore render an accurate account irretrievable.

The experiments reported here are a first step towards understanding the effects of increasing liars' cognitive load on memory for a target event. Nonetheless, there is still much to learn. Most notably, there may be different effects on memory when participants construct their own lies. In the present experiments, some degree of external validity was deliberately sacrificed in favor of isolating the effects of varying cognitive load to test the challenges to the MAD framework outlined in Chapter 3. Nonetheless, since generating information typically enhances memory compared to merely reading it (Bertsch et al., 2007), it is likely that varying cognitive load while participants construct their own lies will moderate the effects on memory. Participants also did not prepare for the interview in advance, but encountered the lies for the first time during the interview itself. Liars typically anticipate and prepare for questions they will be asked (Lancaster et al., 2013) and such rehearsal may also moderate the effect of cognitive load manipulations on memory.

Despite these limitations, the findings presented in this chapter are an important first step to clarifying the independent effects of lie type and cognitive load on memory. Refining our understanding of this relationship will contribute to the development of the MAD framework, as a first theory of lying and memory, and potentially to the refinement of the cognitive approach to lie detection.

Chapter 5: Can fabricating hamper subsequent lie detection in the Concealed Information Test?

5.1. Introduction

In Chapter 1 of the thesis, two reasons were given as to why deception-induced memory impairments could be problematic for the criminal justice system: (1) if an individual retracts a deceptive statement, they may find it harder to retrieve an accurate account of what happened, and 2) it may be harder to detect when somebody is lying if the truth is less accessible in their memory. The findings of Experiments 1-5 support (1) by showing that memory for an event can be less accurate after having lied about that event. Experiment 6 focuses on (2) by investigating whether lying about an event can reduce the accuracy of deception detection later on.

As reviewed in Section 1.5.1, an increasing body of evidence shows that fabricating false information can make the truth less accessible in memory over time (Otgaar & Baker, 2018). Consistent with this research, Experiments 4 and 5 showed that simply reporting fabrications a single time in response to interview questions about a mock crime increased commission errors. Thus, lying in response to interview questions can impair memory for the truth, or make it less salient in memory. A potential consequence of this is that an individual might find lying progressively easier as a criminal investigation proceeds, making it harder for investigators to detect the interviewee's lies in repeated interviews.

Impaired memory for the truth may particularly compromise lie detection when detection strategies rely on a strong memory for the original event. One such strategy is the Concealed Information Test (CIT). By taking various physiological measures of recognition, the CIT determines whether an individual recognizes crime-relevant information that only a guilty person would know. Thus, if fabricating false details early on in an investigation can weaken memory for crime-relevant information, might this lower the accuracy of tests that rely on memory, such as the CIT? This is the question addressed in the final experiment of this thesis.

The CIT is often lauded for its accuracy at detecting when an individual is concealing information, while having a solid theoretical grounding (Meijer, Selle, Elber, & Ben-Shakhar, 2014). In the CIT, examinees are presented with multiple-choice questions concerning the details of the crime. Examinees are shown several possible answers to the question, one of which is the correct crime-relevant detail (the *probe*) and the others are plausible incorrect alternative details (the *irrelevants*). For example, if the suspect stole a necklace, they may be asked “What was the stolen item?” and presented with the possible answers “Ring”, “Brooch”, “Necklace”, “Earrings” and “Bracelet”. The irrelevant details are chosen so that an innocent individual cannot discriminate between them and the probe, resulting in the same physiological response to all answers. A guilty individual, however, recognizes the probe and this elicits a distinct physiological signature termed the “CIT effect” (Lykken, 1974; Verschuere, Ben-Shakhar, & Meijer, 2011).

There are various measures that can be used to assess recognition in the CIT, but the most widely studied are three ANS measures—skin conductance response (SCR), heart rate (HR) and respiration line length (RLL)—and the P300 event-related potential (ERP; (Meijer et al., 2014). The most recent meta-analysis on the CIT’s effectiveness found that all measures yield an impressively large effect size for the difference in physiological responses to probes compared to irrelevants for guilty participants (Meijer et al., 2014).

Because the CIT measures recognition of crime-relevant information, the accuracy of the test is contingent on the quality of the examinee’s memory for that information. If the examinee never encoded or has a weak memory for the item in question, then the CIT effect may be eliminated or attenuated. The majority of the CIT research conducted to date has tested the effectiveness of the test under conditions that elicit strong memory for the crime-relevant information. Researchers often ensure that crime-relevant details are encoded and most conduct the CIT immediately after the event in question (Gronau et al., 2015). In reality, however, examinees might not have encoded all of the crime-relevant details being tested and there can be long delays between the encoding of the event in question and administration of the CIT (Meijer et al., 2014; Osugi, 2011). The examinee may also encounter post-event information in this time period, which could distort memory for important crime-relevant information. Indeed, there are a handful of studies showing that the CIT effect is indeed smaller when memory is compromised in these ways.

Most of the research investigating the effect of compromised memory on the CIT's accuracy have studied how an examinee might deliberately try to manipulate the strength of their memory for crime-relevant information as a way of cheating the test, that is, as a countermeasure. The first experiment to examine this possibility looked at whether voluntarily suppressing memory for crime-relevant details attenuates the P300 ERP component that signals recognition in the CIT (Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013). Participants completed an interactive crime simulation task before undergoing a CIT. Specifically, participants imagined that they were a burglar and 'searched' through a house to steal valuables by pressing numbers on a keyboard that corresponded to different locations in the house. In the CIT participants were shown words describing items from the crime (probes) mixed with words describing items that were not shown in the crime (irrelevants). For some of the probes, participants were instructed to block any memories of the crime from coming to mind when the word appeared on screen. When participants suppressed their memory in this way, the classification accuracy of the CIT fell to an AUC of 0.70, compared to 0.87 when participants were instead instructed to remember as many details as they could about the object when the word appeared. This finding has subsequently been replicated using a similar procedure, which led to indistinguishable P300 components between participants who suppressed their memory for crime-related information and innocents (Hu, Bergström, Bodenhausen, & Rosenfeld, 2015). Together, these studies show that intentional inhibition of the truth can significantly reduce the diagnosticity of the CIT by increasing false innocent classifications.

While intentional suppression of the truth is distinct from memory distortion, additional evidence suggests that the accuracy of the CIT can be compromised regardless of how memory is targeted. One study modified the classic misinformation paradigm to demonstrate that distorting memory for the details of a witnessed crime can reduce the

CIT's diagnosticity. Participants watched a mock crime scenario and—a week later—read a narrative description of the mock crime scenario that contained a number of crime-relevant items. Half of these items were replaced with incorrect misleading information, for instance, a brown envelope was instead described as a red envelope. Participants then completed a CIT to compare their physiological responses for crime-relevant details that were described incorrectly in the narrative text and crime-relevant details that were described correctly (control items). A typical CIT effect was found for control items, that is, a greater SCR response and more pronounced heart rate deceleration. Crucially, there was no CIT effect for crime-relevant details that participants were misinformed about in the narrative text. Thus, misinforming participants about the details of the crime just once meant that they no longer showed physiological signs that they recognized these items. Moreover, participants showed a CIT effect for the misleading information itself. So for the example above, participants showed no physiological response to the correct detail "brown envelope", but instead a physiological response to the incorrect detail "red envelope" (Volz, Bahr, Heinrichs, Vaitl, & Ambach, 2018), which could lead them to be incorrectly classified as innocent.

A crucial difference between lying and misinformation exposure is that people know they are providing false information when they lie, whereas participants in misinformation studies are unaware that they have been exposed to false information. It might be argued that this awareness will make liars more resistant to memory distortion, however existing evidence suggests that the CIT's accuracy can be reduced even when examinees knowingly expose themselves to false information about the crime.

One study examined the effect of deliberately learning false information about a crime as a potential countermeasure for reducing the CIT's diagnosticity by creating memory interference (Gronau et al., 2015). Based on research into retroactive interference, the authors reasoned that learning new false information may interfere with the retrieval of true crime-relevant details that were previously encoded. Participants committed a mock theft and then read a narrative containing false information about some of the key details from the theft they had performed. They were told that learning these false details would help them to appear innocent when they subsequently undertook a CIT. The timing of the exposure to the false narrative and the administration of the CIT were also manipulated to determine if the effect of learning false information on the CIT's diagnosticity is moderated by delays between encoding and test.

CIT detection scores based on SCRs were reduced for participants who learned the false crime details compared to control participants, regardless of when they learnt the false information or took the CIT. However, the overall detection accuracy of the test was significantly reduced only for participants who were exposed to the false crime details and undertook the CIT in a separate session one week after performing the mock theft. These participants also showed impaired recall and recognition of the true crime details, but better memory for the false crime details. The authors therefore propose that the reduced detection accuracy was caused by enhanced memory for the false crime details. Learning the false details immediately before undergoing the CIT meant that memory for these details was stronger than memory for the original details learnt a week before and therefore may have interfered with their retrieval. Recognition of the crime-relevant items was therefore poorer and the CIT less accurate as a result (Gronau et al., 2015).

The handful of existing studies examining how exposure to false information can influence the CIT show that factors known to affect memory do indeed compromise the test's accuracy. This is unsurprising, given that the CIT measures recognition of crime-relevant information and is therefore essentially a physiological memory test. Given the existing evidence that lying can adversely affect memory, the next question is whether merely fabricating information can lead to similar impairments to the CIT. In the real world, people typically generate their own fabrications with the intent to create a false belief in another, but no research has studied whether such fabrication can affect memory and the CIT's diagnosticity accordingly. The present experiment therefore investigated whether self-generating fabricated responses to interview questions about a mock crime can impair memory for those details and affect the diagnosticity of the CIT. In real life, suspects could undergo a CIT weeks or even months into an investigation (Elaad, 2011). It is also possible that they will have been questioned once or on several occasions before it is decided that they ought to undergo a CIT. Given past evidence that exposure to false information can impair the CIT's accuracy, the aim of Experiment 6 was to determine if fabricating false information in interviews prior to a CIT can be a source of misinformation that lowers the test's diagnosticity.

5.2 Experiment 6

The experiment was conducted across three sessions. Participants watched an interactive mock crime and were then questioned on half of the critical items from the crime one week later, where they were asked to lie in response to the questions. Participants were not questioned on the remaining half of the critical items and these therefore served as controls. Two weeks later, participants undertook a physiological CIT that measured SCR and heart rate to assess recognition of crime-related information. The CIT effect refers to a greater SCR response and more pronounced heart rate deceleration to probes compared to irrelevants. It was predicted that participants would show a smaller CIT effect for lied-about items compared to control items.

5.3 Method

5.3.1 Participants.

Eighty participants completed the experiment (mean age = 22.15 years, *SD* = 8.22 years, range = 18-76 years; 27 male, 52 female, 1 other/preferred not to say) in exchange for £10, plus a bonus £20 voucher awarded to the participant with a CIT score most closely approximating an innocent person. Participation was restricted to native English speakers and to individuals who had not participated in any previous studies or pilot experiments conducted for this thesis, due to the overlap in stimuli. To determine an appropriate sample size, a power analysis was conducted using G*Power 3.1.9.2 (Faul et al., 2007) based on the findings of Volz et al. (2018), in which the comparison of interest is participants' responses to crime-relevant items from control categories and crime-relevant items from misled categories. Based on a small effect size of $d = 0.37$, the present experiment required a total of 44 participants to achieve a power of .8 with an alpha level of .05. However, data were collected from 80 participants to compensate for attrition over the three sessions and EDA non-responders (see Section 5.3.2 for definition of non-responders).

5.3.2 Stimuli and apparatus.

Mock crime. This experiment used the mock crime video used in Experiment 4. To recap, the video depicted a gloved individual carrying out numerous actions culminating in the theft of a device and some important documentation. Ten target details were selected from the video for participants to memorize (8 critical, 2 practice). To ensure that participants encoded the details from the mock crime well, the video was edited to make it interactive. Before each of the key details was presented, the video stopped and participants selected the target item required to complete the upcoming action from an inventory that appeared on screen. Because participants watched the mock crime video online, 4 random attention checks were incorporated to ensure that participants were watching the video. Each attention check consisted of a multiple-choice question concerning a detail from the segment of the video that preceded it (details other than target items). Participants selected one of 4 possible answers. The mock crime video was hosted on Qualtrics (Qualtrics, Provo, UT). All video controls were disabled so that participants could not rewind, fast forward, pause or replay the video.

Physiological recording. Two physiological measures were recorded for the CIT: Skin conductance response (SCR) to assess orienting and heart rate (HR) to assess arousal inhibition. Orienting and arousal inhibition are the primary mechanisms known to underlie the CIT effect. The orienting response is a reaction to significant or novel stimuli. Stimuli that are significant to the individual—such as crime-relevant information—elicit an orienting response; that is, enhanced SCRs (Klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2017). Arousal inhibition refers to inhibition of the experienced physiological arousal in response to crime-relevant information, which manifests as more pronounced HR deceleration (Klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2016).

Physiological data were logged using a MP36R data acquisition unit (*Biopac Systems Inc*) with pre-gelled disposable Ag/AgCL electrodes (EL507 for SCR and EL501 for HR). Participants were also video recorded to identify any noise generated from excessive movements and to ensure that these movements did not overlap with stimulus presentation. Data were excluded for any trial where there was an excessive movement up to 2s before stimulus onset (Klein Selle et al., 2016).

To record SCR, electrodes were placed on the distal phalanges of the first and middle finger of the non-dominant hand. SCR peaks were identified by the AcqKnowledge v4.2 proprietary algorithm (Kim, Bang, & Kim, 2004) and were sampled at 1000Hz at x2000 gain with a 66.5Hz low pass filter. Parameters were set so that peak onsets were within 0.5-5s following stimulus onset and maximum peaks within 10s (Gamer, 2011). Peaks identified by the algorithm that were not linked to stimulus presentation were removed. The skin conductance response analyzed was the difference in the absolute magnitude of tonic skin conductance peaks and their peak onsets. Approximately 25% of the general population do not elicit valid SCRs and are considered 'non-responders' (Venables & Mitchell, 1996). Participants were considered non-responders if the standard deviation of their raw SCRs was below 0.01μS and their SCR data were excluded from analyses (Klein Selle et al., 2016). SCRs for the first item in each CIT block were excluded from analysis, as this item was a buffer item to absorb the initial orienting response (Gershon Ben-Shakhar & Elaad, 2002).

Heart rate was recorded from two electrocardiogram (ECG) electrodes placed on the ventral side of the non-dominant wrist and the non-dominant lateral aspect of the distal fibula, and from the SCR electrode located on the non-dominant distal phalange. ECG signals were sampled at 1000Hz at ×1000 gain, with a 66.5Hz low pass filter and a 0.5Hz high pass filter. To assess the change in heart rate following stimulus presentation, the ECG signal was processed using *AcqKnowledge*'s propriety Heart Rate Variability algorithm to detect R peaks, measure the time interval between R peaks and filter artifacts. The algorithm identifies the change in heart rate by converting the R-R time interval to beats per minute (bpm) and performing baseline-correction by subtracting the mean heart rate in the 3s before stimulus onset. This is compared to the mean baseline-corrected heart rate in the 15s following stimulus onset to yield the change in heart rate following stimulus presentation (Gamer, 2011).

To reduce noise from the variability in the magnitude of physiological responses across participants, all physiological measures were normalized by converting them to z-scores and these z-scores were used as the units for analysis (Ben-Shakhar, 1985).

5.3.3 Design and procedure.

The experiment was a 2 (Veracity: Lie, Control) x 2 (Item Type: Crime-relevant, Crime-irrelevant) within-subjects design. The experiment consisted of three sessions completed over 3 weeks. Participants watched the mock crime in Session 1 and were questioned about the crime in Session 2, which took place 1 week after Session 1. The mock crime and questioning phases of the experiment were separated in line with Gronau et al. (2015) and Volz et al. (2018), who both introduced misinformation in a separate session to the mock crime. Past research also shows that fabrication has the largest effects on memory when lies are constructed in a separate session to the original event (Polage, 2012). Finally, participants completed the CIT in Session 3, which took place 2 weeks after Session 2. Sessions 1 and 2 were completed online, whereas Session 3 took place in the laboratory. Participants were given a cover story to conceal the memory aspect of the experiment. They were told that the experiment concerned the performance of the CIT at detecting plausible lies after a time delay.

Session 1. After consenting to participate and providing demographic information, participants read a background story to immerse themselves in the experiment. They were asked to imagine that they had found themselves in financial difficulty and had resorted to a loan shark to cover their expenses. To repay their loan, they were to steal a valuable device and the blueprints detailing how the device is made. Participants then received instructions for how to carry out the crime, which consisted of 10 statements, each outlining an action together with the item needed to complete the action written in red (e.g. “lever open the workshop door using a crowbar”). Participants were instructed to memorize the items in red and had 90 seconds to do so. The first and last items were filler/practice items to control for primacy and recency effects, leaving 8 critical items for memorization.

Participants then ‘performed’ the mock crime by watching the interactive mock crime video, consisting of the 10 actions outlined in the instructions, and were asked to imagine that they were the person in the video performing the actions shown. Immediately before each action in the video was executed, an inventory appeared on the screen showing the 10 items and a description of the upcoming action (see Figure 5.1 for examples). Participants selected the item required to complete the action and the mock crime video then resumed.

The inventory items were arranged in a random order, but the order was kept constant across all 10 actions. In addition, participants were interrupted with a multiple-choice attention check question at 4 pre-determined random intervals. On completion of the mock crime, participants were asked if they had seen the mock crime video before in any previous experiments they had participated in (none had). Participants were then thanked and reminded that they must complete Session 2 of the experiment in one week’s time.

Session 2. Participants were emailed a link to Session 2 of the experiment 7 days after completion of Session 1 (mean delay between sessions = 7.12 days, $SD = 0.43$). Participants were questioned about 4 of the 8 critical details from the mock crime. The remaining 4 critical items served as control items that were not addressed during questioning. Each of the 8 items were randomly allocated to one of two sets and half of the participants were questioned about the items from Set 1 and the remaining half were questioned about the items from Set 2. The set that participants completed was randomly determined by Qualtrics.

Before questioning began, participants were reminded of the mock crime scenario from Session 1 and were asked to imagine that they had been arrested. They were informed that the police had CCTV evidence showing them at the crime scene at the time the crime was committed and that the police knew that they were guilty, but were missing 4 key pieces of information that they needed to locate the device. Participants were informed that they would be interviewed about this information, but that they should avoid incriminating themselves further by lying in response to all questions by creating a plausible alternative answer. To encourage participants to take questioning seriously and to reinforce the cover story, participants were told that the plausibility of their lies would be assessed.

Participants were instructed to give 1-2 sentence answers and had 1 minute per question to think of and type their response. They were first shown an example question and answer to demonstrate the type of response that was expected (see Figure 5.2). Participants then completed a practice question, typed their response, and were shown an example of how they could have responded to ensure that they understood the task. They then proceeded to answer 4 questions addressing each of the critical mock crime details (for example, "What did you use to cut the padlock off the device case? Please briefly describe what the item you used looked like"). The critical questions were presented in a random order across participants. Participants typed their response and moved onto the next question. On completion of all questions, participants were shown their answers to each question and were asked to confirm that the answers shown were the ones they provided. This was to reinforce their lies in memory. Finally, participants were reminded that they must come to the lab in 2 weeks time to complete a lie detection test.

Session 3. Participants were invited to the laboratory to complete the CIT 14 days after completion of Session 2 (mean delay between sessions = 14.21 days, *SD* = 0.53). They were informed that they were to undergo a physiological lie detection test concerning the items from the crime and that the person with a CIT score most closely approximating an innocent person would win a £20 Amazon voucher.

The CIT consisted of 9 blocks, 1 practice block and 8 critical blocks (one for each critical item from the mock crime). Each block contained a probe (a critical item from the mock crime) and 4 irrelevants (alternative items not shown in the mock crime). If a participant had provided one of the irrelevants as their lie during questioning in Session 2, that irrelevant was substituted for an alternative irrelevant item in that CIT block. Participants' responses to the questions in Session 2 were reviewed so that each participant's CIT could be customized to swap irrelevants where necessary (participants required the substitute irrelevant for a mean of 19.20% [*SD* = 39.44%] of items). Appendix 10 lists the CIT questions, probes and irrelevants for each critical item. To ensure that the irrelevants were unbiased alternatives to the probes, 20 additional 'innocent' participants who had not seen the mock crime video performed the CIT. Paired sample t-tests were performed to compare physiological responses to the probe and irrelevants for each item. There were no significant differences in responses to probes and irrelevants for any item in neither EDA nor HR measures (all *p*s <.05).

The CIT was conducted in the laboratory with the experimenter present in the room at all times (located behind a screen for the duration of the CIT). Participants were connected to the electrodes and sat at a desk approximately 1 meter from a 22-inch monitor with a resolution of 1920x1080. They were informed that a series of questions about the mock crime from Session 1 would appear on the screen followed by a series of possible answers, presented one-by-one.

The task was to state "no", aloud, in response to every answer that appeared onscreen. The first CIT block was a practice block addressing a filler item from the mock crime to familiarize participants with the task. On completing the practice block, participants were given the opportunity to ask any questions before starting the main CIT. For each of the 8 critical CIT blocks, a question appeared in the center of the screen for 10 seconds followed by the 5 possible answers, sequentially presented as text on the screen for 5 seconds. The inter-stimulus interval between each answer was 10 seconds. The first answer presented was always an irrelevant to serve as a buffer item to absorb the initial orienting response.

On completing the CIT, participants completed a recognition test to assess their explicit memory for the 8 critical items from the mock crime. Participants were informed that the lie detection phase of the experiment was over and that they should answer all questions truthfully according to their memory of what was shown in the mock crime in Session 1. The recognition test consisted of 8 multiple choice questions addressing each critical item from the crime with 5 possible answers: the true answer and the 4 irrelevants for that item from the CIT. The order of the questions and the 5 answers for each question was randomized across participants. After each recognition question, participants rated their confidence that their answer was correct on a scale of 0 = *completely uncertain* to 100 = *completely certain*.

Figure 5.1. Screen shots extracted from the interactive mock crime video depicting two imminent actions (left panels). The video paused immediately before each action was performed and participants selected the correct item to execute the action from the inventory (right panels).

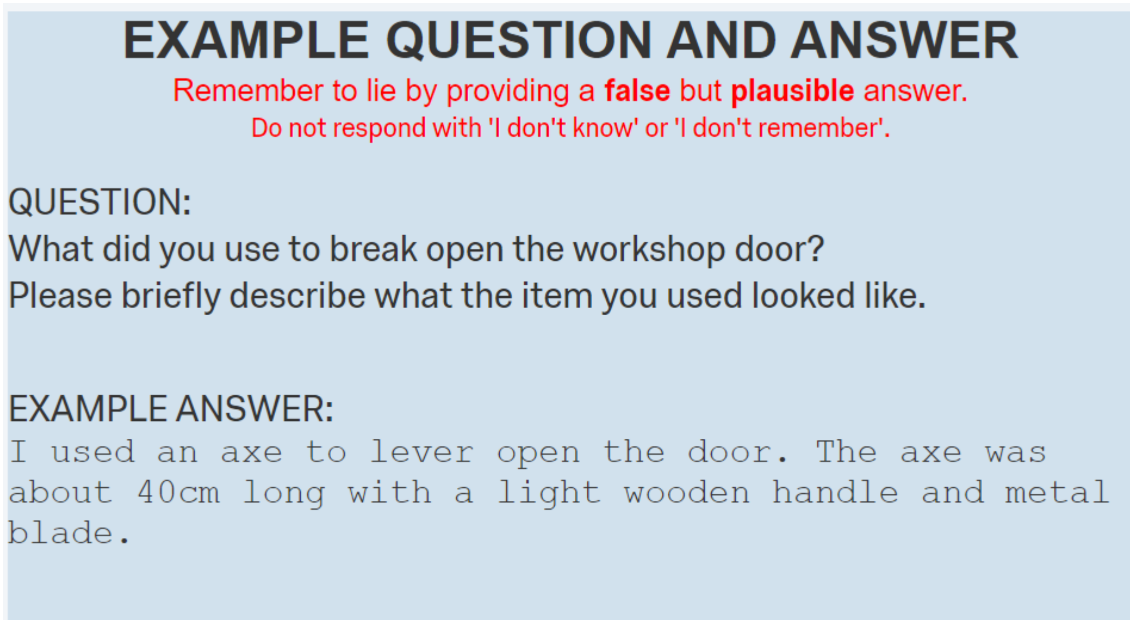


Figure 5.2. Example question and answer provided to participants to demonstrate how they were expected to respond to questions.

5.4 Results

5.4.1 Preliminary analyses

Before addressing the primary research question of the experiment, three performance checks were performed.

Engagement in mock crime. Participants' performance in the interactive mock crime and the 4 attention checks were analyzed to assess whether participants were sufficiently engaged with the mock crime. Participants passed an average of 86.88% (*SD* = 33.82%) of attention checks, indicating that they paid attention to the mock crime video. Additionally, participants selected the correct item to perform each action described for 98.75% (*SD* = 11.12%) of items, suggesting that they encoded the crime-relevant items well.

Performance during questioning. Of the 80 participants who completed Session 1 of the experiment, 75 completed Session 2. The responses that participants provided during questioning were examined to ensure that they followed instructions by providing a false alternative answer. Participants answered correctly in an average of 96.27% (SD = 19.98%) of questions. When participants answered incorrectly (by providing the truthful answer) or did not provide an answer, the item in question was excluded from further analyses.

Explicit memory performance. Of the 75 participants who completed Session 2, 71 completed Session 3. Analysis of participants' performance in the final recognition test indicated that lying during questioning did not impair explicit recognition of lied-about items: There was no significant reduction in the proportion of items that participants correctly recognized from the mock crime for lied-about (M = 0.80, SD = 0.20) compared to control items (M = 0.78, SD = 0.24, t(70) = -0.57, p = .57). In fact, lying during questioning may have slightly improved memory, as suggested by significantly higher confidence ratings for lied-about (M = 80.85, SD = 12.83) compared to control items (M = 77.48, SD = 13.31, t(70) = -2.05, p = .04, d = 0.26). Self-generating fabrications in response to questions about the mock crime therefore did not adversely affect explicit memory for the critical mock crime items.

5.4.2 Physiology

Skin conductance. Nine participants were flagged as non-responders (as described in Section 5.3.2) and were excluded from analysis, leaving SCR data from 62 participants for analysis. This still well exceeds the target sample size of 44 participants.

It was predicted that participants would show smaller SCRs in response to items they lied about during questioning in Session 2 compared to control items that they did not lie about. To test this hypothesis, a 2(Item Type: probe, irrelevant) x 2(Veracity: control, lie) repeated-measures ANOVA was performed on the normalized SCR scores. This revealed a significant main effect of Item Type (F(1,61) = 18.30, p < .001), which represents the CIT effect. As shown in Figure 5.3, SCRs were larger for probes (critical items shown in the mock crime; M = 0.21, SD = 0.55) compared to irrelevants (alternatives not shown in the mock crime; M = -0.07 , SD = 0.28).

There was no main effect of Veracity ($F(1, 61) = 0.97$, $p = .33$) and no significant interaction between Item Type and Veracity ($F(1,61) = 0.04$, $p = .84$), indicating that lying in response to the questions in Session 2 did not affect SCRs. In sum, participants showed larger SCRs in response to items from the mock crime, regardless of whether they lied about those items during questioning. Thus, contrary to the hypothesis, lying during questioning did not affect SCRs in the CIT.

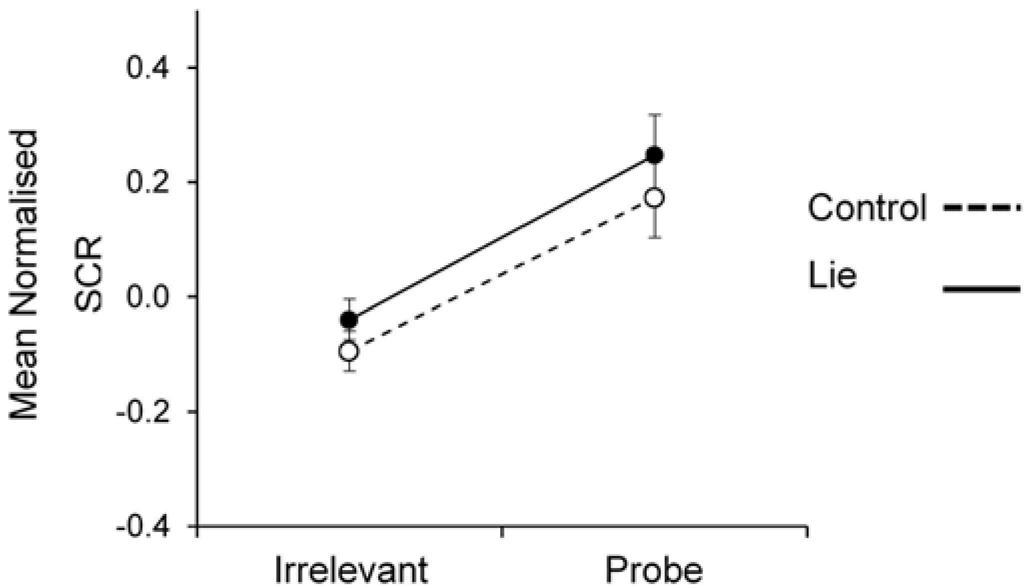


Figure 5.3. Mean normalized SCR scores for probe and irrelevant items for mock crime items that participants lied about during questioning and mock crime items that participants were not asked about during questioning (control). Error bars represent ±1 SE of the mean.

Heart rate. Heart rate deceleration values were analyzed from the 71 participants who completed all three study sessions. It was predicted that participants' HR would be lower in response to items that they lied about during questioning in Session 2 compared to control items that they did not lie about. To test this hypothesis, 2(Item Type: probe, irrelevant) x 2(Veracity: control, lie) repeated-measures ANOVA was performed on the normalized HR deceleration values.

The HR data followed the same pattern of results as the SCR data. The analysis revealed a main effect of Item Type ($F(1,70) = 13.39$, $p < .001$), which represents the CIT effect. As shown in Figure 5.4, participants' HRs were lower for probes ($M = -0.12$, $SD = 0.48$) compared to irrelevants ($M = 0.05$, $SD = 0.23$).

There was no main effect of Veracity ($F(1,70) = 0.01$, $p = .92$) and no significant interaction between Item Type and Veracity ($F(1,70) = 0.02$, $p = .89$), indicating that heart rate deceleration was unaffected by lying during questioning. Thus, contrary to the hypothesis, lying during questioning did not affect HR deceleration—participants showed a CIT effect regardless of whether they lied during questioning.

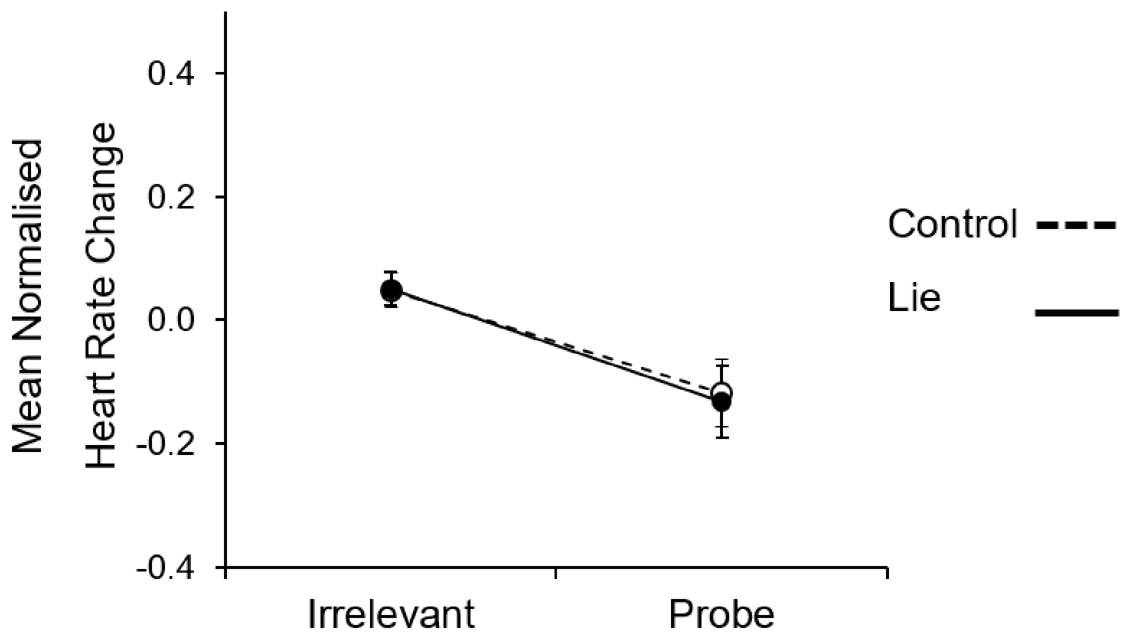


Figure 5.4. Mean normalized heart rate deceleration for probe and irrelevant items for mock crime items that participants lied about during questioning and mock crime items that participants were not asked about during questioning (control). Error bars represent ±1 SE of the mean.

5.4.3 Signal detection analysis

ROC curves were plotted to assess the detection efficiency of the CIT. Detection efficiency is calculated by comparing the overlap between the distributions of guilty and innocent participants' scores and can range from 0-1. A value of 1 indicates that there is no overlap between the distributions and therefore represents perfect detection, whereas a value of 0.5 indicates that the distributions of guilty and innocent participants cannot be distinguished and therefore represents chance detection. A ROC curve can be created from these distributions and the area under the curve (AUC) tells us the classification accuracy of the test (Gershon Ben-Shakhar & Elaad, 2003).

Because all participants watched the mock crime and were therefore "guilty" a group of innocent participants of equal sample size was simulated. Simulation of innocents involves computing values for the probe and irrelevant items that do not significantly differ from one another. This represents an innocent individual who has no knowledge of the crime items and therefore should not respond differently to probes compared to irrelevants. Simulation of innocents is commonly conducted in the CIT literature (e.g., (Klein Selle et al., 2016; Meijer et al., 2014) and the standard method is to randomly draw an SCR value for each item from a standard normal distribution with a mean of 0 and SD of 1. This is calculated for each participant and the resulting values are used to create the ROC curves.

The ROC curve plots the true positive rate (i.e., sensitivity: Participants classified as guilty who were in fact guilty) as a function of the false positive rate (i.e., 100-specificity: Participants classified as guilty who were in fact innocent) for the range of SCR values. An ideal test would show a high true positive rate, while maintaining a low false positive rate, and therefore would show an ROC curve that lies as close to the co-ordinate (0,1) as possible and an AUC close to 1 (DeLong, DeLong, & Clarke-Pearson, 1988). Figure 5.5 shows that the ROC curves for both lied-about and control items lie close to the diagonal chance line, indicating that the CIT did not discriminate between guilty and innocent participants well in this experiment.

Pairwise comparison of ROC curves for control and lied-about items was performed using MedCalc (MedCalc Software, Ostend, Belgium), which uses a non-parametric Mann-Whitney U-Statistic to compare within-subject ROC curves (as recommended by DeLong, DeLong, and Clarke-Pearson, 1988). This revealed no significant difference in

AUCs for control and lied-about items (AUC_{Diff} = .05, 95%CI[-.08, .19], *p* = .44), indicating that lying during questioning did not reduce the diagnosticity of the CIT for discriminating between guilty and innocent participants. Nonetheless, while the detection efficiency of the CIT for control items was significantly greater than chance (*AUC* = .63, 95% CI[.54, .71], *p* = .01), the detection efficiency for lied-about items did not significantly differ from chance (*AUC* = .57, 95%CI[.48, .67], *p* = .44). This suggests that there might have been a slight reduction in the CIT's performance for items that participants lied about during questioning. It should be noted, however, that detection efficiency was poor regardless of whether participants lied; AUCs for both control and lied-about items were below the acceptable rate of discrimination of .70-.80 (Hosmer & Lemeshow, 2000).

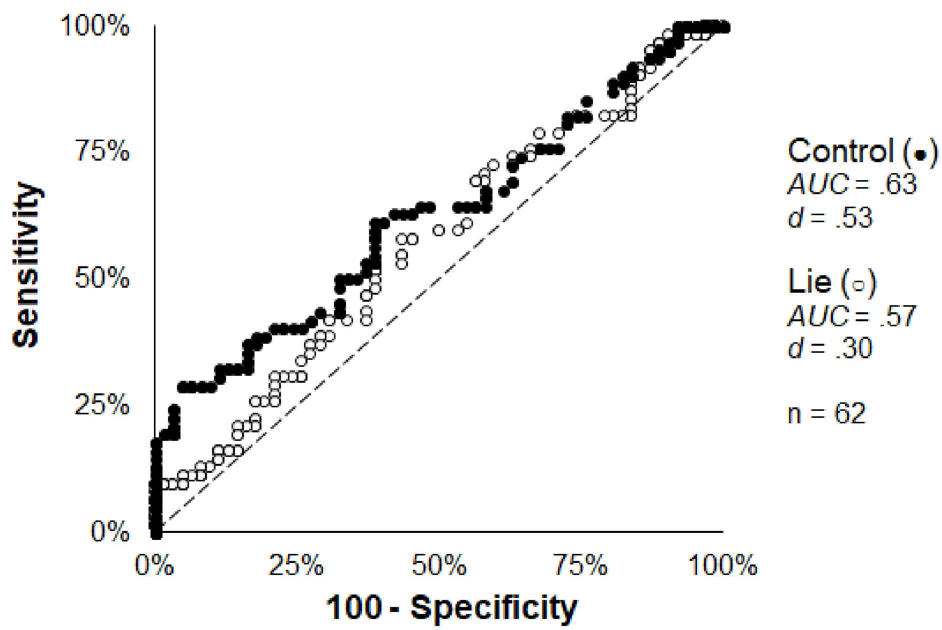


Figure 5.5. ROC curves showing the signal detection rate of the CIT effect between guilty and simulated innocent participants for items that participants did not lie about during questioning (Control) and items that participants lied about during questioning (Lie). The dotted line represents chance detection. *AUC* = area under the curve, *d* =

Cohen's *d* indicates the effect size of the difference between guilty and simulated innocent participants.

5.5 Discussion

While most research shows impressive accuracy for the CIT at detecting concealed knowledge, the studies reviewed in Section 5.1 uniformly demonstrate that the test's accuracy depends on the quality of the examinee's memory for the event in question. This is unsurprising, given that the CIT measures recognition of crime-relevant information. It is therefore important to establish what conditions impair memory for crime-relevant information, since this will inevitably compromise the CIT's accuracy. The present experiment showed that self-generating fabrications in response to interview questions about a mock crime experienced 3 weeks before was insufficient to impair memory. Accordingly, the CIT effect was unaffected for items that participants lied about (compared to control items) for both SCR and HR measures. Nonetheless, the ROC analysis indicated that the detection efficiency of the CIT did not significantly differ from chance for lied-about items, but was significantly above chance for control items. This suggests that lying during interview may have slightly impaired the accuracy of the CIT at distinguishing guilty from simulated innocent participants. However, this finding should be interpreted with caution, given that the detection accuracy did not significantly differ between control and lied-about items.

Previous research most closely resembling the present experiment are studies conducted by Volz et al. (2018) and Gronau et al., (2015), who both demonstrated that exposing participants to false information prior to a CIT significantly impaired memory and the test's diagnosticity. One notable difference between those studies and the present is that participants in this experiment were explicitly instructed to lie by generating their own false information, whereas participants in previous research were exposed to or learnt experimenter-generated false information. It is therefore possible that self-generating fabrications protected participants from any memory distortion that might compromise the CIT.

Why might self-generating fabrications have protected participants? It is well established that people remember information better when it is self-generated rather than merely read—the *generation effect* (Slamecka & Graf, 1978). In the present context, the generation effect may have improved participants' memories for the lies that they generated in Session 2 compared to if they had been provided with ready-made fabrications. According to the *lexical activation hypothesis*, semantic features of an item are activated in memory when that item is self-generated as opposed to merely read. These semantic features are thought to provide retrieval cues when the individual later attempts to recall the item, thus facilitating recall (Payne, Neely, & Burns, 1986). Source memory may also benefit from self-generation, particularly when participants distinguish internally generated information from externally provided information, as in the present case (Riefer, Chien, & Reimer, 2007). Self-generating fabrications may therefore have improved participants item and source memory for their lies, making it easier for them to later distinguish fabricated details from those originally experienced in the mock crime and thereby protecting them from memory distortion. Future research should directly compare the effects of self-generated and experimenter-generated fabrications on subsequent memory to determine if generation effects are responsible for protecting participants against deception-induced memory distortion (although generation effects are difficult to isolate in a deception context, given the variability in the lies that participants generate).

While it is possible that the generation effect prevented deception-induced memory distortion by improving memory for the lies told, the opposite may also be true; that is, that participants' memories were unimpaired because they had *poor* memories for the lies told. Since participants reported their lies only once, they may not have encoded them sufficiently to create source monitoring difficulties at test. While participants in Volz et al. (2018) also encountered false information once, they were unaware that they were exposed to this information and such naivety can increase vulnerability to misinformation effects (Blank & Launay, 2014). When participants knowingly learnt false information as a countermeasure in Gronau et al. (2015), they rehearsed it until they achieved perfect performance. As outlined in Section 5.1, this likely created a strong memory for the false information that blocked retrieval of the original event details via retroactive interference. Fabricating a false answer only once may not have created a strong enough memory to cause this interference at retrieval. Thus, if retroactive interference is indeed the mechanism underlying the memorial effects that hampered the CIT in Gronau et al., it may be that the conditions in the present experiment were insufficient to create interference and consequently no impairment to the CIT was found. It was a deliberate choice to request that participants lie only once, as past research shows that repeating self-generated fabrications can decrease the likelihood of deception-induced memory impairment (Rindal, 2017). Nonetheless, it is possible that repeating lies over time might affect memory differently and the CIT's effectiveness accordingly.

While the present experiment showed no deception-specific impairment to the CIT, it should be noted that the detection efficiency of the test was poor overall, with an AUC of 0.63 and an effect size of $d = 0.57$ for the difference in responses to the probes for guilty and innocent participants. This is markedly smaller than the average effect size of $d = 1.55$ found in CIT studies that typically implement optimal conditions for encoding crime-relevant information (Meijer et al., 2014). Indeed, the meta-analysis conducted by Meijer et al. excluded studies with a delay between encoding of the event in question and the administration of the CIT. However, the results of the present study suggest that a delay of just 3 weeks may be enough to considerably decrease the difference in physiological responses to crime-relevant items between guilty and innocent examinees. Participants' explicit memory performance for control items was just 78%, suggesting that they forgot a significant proportion of critical items from the mock crime in interval between encoding and test. Thus, this experiment adds to a handful of previous studies suggesting that even relatively short time delays may be problematic for the diagnosticity of the test in the field (Gamer, Kosiol, & Vossel, 2010; Nahari & Ben-Shakhar, 2011; Seymour & Fraynt, 2009).

In summary, this experiment provides preliminary evidence that self-generating fabrications in a single interview is not sufficient to affect memory and the CIT's accuracy after a 3-week delay. Nonetheless, given the increasing body of research showing that lying about an event can impair memory for that event, further research is warranted to examine how lying under different circumstances might affect memory and consequently the accuracy of the test.

Chapter 6: General Discussion

This thesis began by noting the prevalence of deception at every level of the criminal justice system. Given its prevalence, there has been surprisingly little investigation into how lying can affect not only the beliefs of the deceived, but also those of the deceiver. The research presented throughout adds to the existing small body of evidence showing that by deceiving others, people can impair their ability to know the truth themselves.

Although the effect of lying on memory has received increasing attention from applied memory researchers, there has been little theoretical development in this area of research and particularly little discussion of the mechanisms by which lying may distort memory. As such, this thesis was built around the two relevant theoretical frameworks of the relationship between lying and memory: Von Hippel and Trivers' (2011) theory of self-deception and Otgaar and Baker's (2018) MAD framework. The main aim of this thesis was to contribute to theory development to enable a better understanding of when deception-induced memory distortion could be problematic for the functioning of the criminal justice system. The experiments conducted therefore examined the mechanisms that allegedly underlie deception-induced memory distortion, as proposed by von Hippel and Trivers and the MAD framework, as well as the potential impact of such memory impairments on deception detection. Given that each of the previous chapters contain a general discussion, this chapter will bring together the theoretical and practical conclusions that can be drawn from the experiments presented within this thesis, as well as consider their limitations and avenues for future research.

6.1. Theoretical implications

The first three experiments of the thesis tested von Hippel and Trivers (2011) proposal that self-deception might be achieved via RIF of the truth. To recap, von Hippel and Trivers state that simply repeating a lie and omitting the truth may promote RIF of the omitted information, thereby helping people to lie convincingly and escape detection. As outlined in Section 1.4, we must be able to identify "below baseline" forgetting to differentiate RIF from mere decay over time. However, prior to the experiments conducted in Chapter 2, no existing research had implemented the appropriate design and controls to determine that RIF is indeed the mechanism underlying forgetting effects.

Section 2.1 outlined the parallels between the activation-competition-inhibition cycle known to underlie RIF in broader memory research and the processes known to be involved when people fabricate. It was therefore proposed that repeating fabrications might indeed trigger inhibition of the truth and reduce its accessibility. Experiments 1-3 were therefore the first to test if repeating self-generated fabrications can promote RIF of the corresponding truthful information via an inhibitory mechanism. No evidence was found for this hypothesis: Participants' memories for lied-about items did not differ from baseline memory. Repeating fabrications under the conditions tested was therefore equivalent to not rehearsing the items between encoding and recall, suggesting that any reduction in memory performance was due to decay over time and not caused by repeating fabrications *per se*. Thus, there is still no adequate evidence for von Hippel and Triver's claim that self-deception can be achieved by RIF.

Nonetheless, Experiments 1-3 implemented very constrained adaptations of the original RIF paradigm (Anderson et al., 1994) by matching its conditions as closely as possible. It may be that repeating lies and omitting the truth does promote RIF of the truth when lies are repeated over time, for instance. However, implementing a time delay may obscure RIF effects with source monitoring difficulties. Indeed, the experiments outlined in Chapter 4 add to the existing body of evidence that source monitoring plays a crucial role in deception-induced memory distortion.

In Chapters 3 and 4, the primary focus was on the relationship between cognitive load and source monitoring errors when people fabricate. Cognitive load was considered important to investigate, as it is one of the key factors that distinguishes lying from other forms of misinformation and because it plays an important explanatory role in both the MAD framework and von Hippel and Trivers' theory of self-deception. Accordingly, Chapters 3 and 4 focused on the MAD framework's proposals concerning the relationship between the cognitive effort of lying and subsequent memory errors.

In Chapter 3, it was proposed that cognitive load might not be an appropriate primary predictor of the memory errors that follow deception and therefore that the framework may require a shift of focus from cognitive load to other causal mechanisms; namely, a lack of rehearsal in the case of false denial and feigned amnesia, and source monitoring errors in the case of fabrication. Specifically, it was proposed that the cognitive effort of lying and the type of lie should be considered independent variables and that failing to dissociate them could obscure the relationship between the cognitive effort of lying and subsequent memory errors. While the MAD framework proposes that increasing the cognitive effort of lying should increase commission errors, past research outlined in Chapter 3 suggested that if we isolate fabrication, commission errors may in fact be highest when participants fabricate under low cognitive load and lowest when fabricating under high cognitive load.

The experiments reported in Chapter 4 proceeded to test the above hypothesis. While this hypothesis was not fully supported, the findings of Experiment 5 provided evidence that the cognitive effort of lying and lie type can dissociate and interact to affect memory differently. Varying the cognitive effort of a single type of lie—fabrication—led to different memorial effects depending on how cognitive load was manipulated.

The findings of Experiment 5 provide evidence that the MAD framework may benefit from explicitly separating the type of lie from other factors that might moderate the effects of lying on memory, including cognitive effort. The findings of Experiments 4 and 5 suggest that the question of whether lying increases commission errors seems to depend on whether there are counterfactual details in memory that must be discerned from truthful details: Experiment 5 showed that commission errors increased only when participants encoded their lies. In other words, source monitoring is likely the most important factor for predicting commission errors. This is the most parsimonious explanation of when lying promotes commission errors; one cannot falsely recall a detail that was never encoded.

The most appropriate next line of enquiry may therefore be to determine the moderators and mitigators of deception-induced commission errors. Factors known to affect source monitoring ability, such as those discussed in Chapter 1, will likely moderate the rate of commission errors that lying produces. Thus, lying may not have unique effects on memory compared to a standard misinformation effect. Knowledge of the fact that the information communicated is false might not be sufficient to protect people from memory errors; wherever source needs to be monitored, errors inevitably creep in.

Predicting commission errors based on what we know to affect source monitoring may lead to more accurate predictions than predicting commission errors from the cognitive effort of lying. For instance, plausible fabrications require less cognitive effort to construct than implausible fabrications (Walczyk et al., 2014), yet we know that plausible lies also promote commission errors (Pickel, 2004). The source monitoring framework is consistent with this finding—we know that the plausibility of a false event can affect the likelihood and extent of impaired source monitoring (Hart & Schooler, 2006; Pezdek, Blandon-Gitlin, Lam, Hart, & Schooler, 2006; Sharman & Scoboria, 2009) and we would therefore predict an increase in commission errors for plausible lies. However, if we instead based our prediction on the cognitive effort of lying, we would incorrectly predict fewer commission errors for plausible lies.

In sum, the findings across this thesis do not support the idea that RIF can underlie self-deception. However, they do support the proposition that we can experience "self-inducing false memories" (von Hippel & Trivers, 2011, p. 6) via source monitoring errors. Moreover, we can alter the likelihood that commission errors will increase after fabrication by manipulating cognitive load. Nonetheless, the cognitive effort of lying should be considered a potential moderator of commission errors, rather than the primary predictor.

6.2 Practical implications

In real criminal cases, we might worry about the effect of lying on memory when an individual changes their testimony in an investigation; for instance, if they retract a deceptive testimony and revert to truth telling. This may be particularly important in cases of witness intimidation, where an eyewitness to a crime is intimidated or bribed into providing a false testimony or alibi. In 2016, an alleged instance of this brought significant media attention. David Osadebay, a well-known rapper and gang member, was accused of plotting a drive-by shooting of Oliver Tetlow, a bystander mistaken for a rival gang member. The case looked like a certain conviction until a witness suddenly changed his story during the trial, providing Osadebay with the alibi that led to his acquittal. It is suspected that this alibi was in fact false and that the witness was intimidated into fabricating a false account to weaken the prosecution's case (Haydock, 2018; Wright, 2017).

This case, like many others, highlights the importance of understanding how previous fabrications might affect an individual's ability to later recall an event accurately. Experiments 1-5 add to the existing body of evidence showing that fabricating such false details could indeed impair the witness' memory, potentially rendering a fully accurate account irretrievable if they later reverted to truth telling. Experiment 2 demonstrated that fabricating information impaired memory for information that was related to the lied-about details by 10%. Thus, if such information became relevant later in an investigation, an individual may be less likely to recall those details accurately because of having initially lied about *other* related details. Although only a small impairment, seemingly insignificant single details can make or break a criminal case (Van Duisen, 2014). Thus, any factor that compromises memory, however little, should be taken seriously.

The memory impairments seen in Experiments 4 and 5 were somewhat larger: Participants who lied in Experiment 4 showed a relative increase in commission errors of 50% and 38% when they lied under low and high cognitive load respectively, compared to participants who did not lie (though the absolute increase was 8% and 6% respectively). Similarly, commission errors doubled when participants lied under low cognitive load in Experiment 5 (with an absolute increase of 19%) and increased by 55% when participants lied under high intrinsic load (with an absolute increase of 12%). Taken together, these findings demonstrate that the effect of fabrication on memory for the truth is a legitimate concern for revised testimonies.

The next question, then, is how might we mitigate such errors? Given that we cannot stop people from lying, one approach might be to consider questioning approaches that reduce the chance that lying will contaminate memory. The findings of Experiment 5 suggested that one potential way to protect interviewees from deception-induced memory distortion could be to ask questions that target source memory, rather than just item memory: When participants were questioned under high intrinsic load, commission errors increased when item memory was targeted, but source memory performance remained high. This finding is consistent with those of broader memory research showing that directing people to consider the source of their memories can reduce false memory rates by encouraging them to apply stricter source monitoring criteria (Dodson, Koutstaal, & Schacter, 2000; Lindsay & Johnson, 1989). Thus, it may be wise to obtain a revised witness testimony by encouraging witnesses to consider the source of their memory to yield the best chances that their testimony is accurate. Further research should investigate this possibility.

Not only might deception-induced memory distortion create problems for revised testimonies, it might also compromise deception detection. Von Hippel and Trivers (2011) proposed that one of the ways to reduce the cognitive cost of lying—and therefore increase the chances of remaining undetected—is to make the truth less salient in memory. Reducing the accessibility of the truth means that less effort is expended on processes such as monitoring the behavior of the enquirer and the self, as well as suppression of the truth, thereby reducing potential cues to deception that might otherwise arouse suspicion. This is the reasoning behind the cognitive approach to lie detection: Deceivers implement strategies to reduce the cognitive cost of lying (von Hippel & Trivers, 2011; Walczyk et al., 2014), but imposing cognitive load counters this attempt, thereby magnifying cues that are typically faint and unreliable (Vrij, Fisher, et al., 2008).

While there has been some disagreement over the effectiveness of the cognitive approach (Levine, Pete Blair, & Carpenter, 2018), the consensus is that imposing cognitive load on interviewees successfully improves deception detection with a medium effect size (Vrij, Blank, & Fisher, 2018). The findings of Experiment 5 may explain one way in which this approach magnifies cues to deception like inconsistencies: Dividing participants' attention reduced their encoding of the lies reported during questioning, which could promote inconsistencies over repeated interviews and therefore help to expose deception (Vredeveldt et al., 2014). Importantly, the findings of Experiment 5 suggest that dividing attention should selectively promote inconsistencies in liars, but not truth tellers: Participants' memory was not adversely affected when they told the truth. A possible avenue for future research is to consider how dividing liars' attention across multiple interviews might increase inconsistencies and facilitate detection.

Finally, Experiment 6 directly tested the hypothesis that the mere act of fabricating in an initial interview might hamper an objective deception detection test, the CIT. Given that past research indicates that lying just once can adversely affect memory and that the CIT relies on a strong memory for the truth, it is possible that previously fabricating might lower the accuracy of the CIT. However, Experiment 6 did not find evidence for this: Fabricating details about a mock crime was insufficient to affect memory for the details originally witnessed or to affect the accuracy of the CIT. Thus, eluding detection via memory distortion of the lied-about information likely requires that deliberate strategies are implemented to achieve this, such as intentional suppression of the truth (as in (Bergström et al., 2013), or studying false information (as in Gronau et al., 2015). For deception-induced memory distortion to be a more passive process, it may instead require repetitions of the fabricated information over extended periods of time. An interesting question for future research regarding von Hippel and Trivers' (2011) theory is whether manipulating the motivation to elude detection increases people's susceptibility to memory distortion and ability to 'pass' the CIT. This would provide evidence that memory distortion is indeed a mechanism to better deceive others.

It should be highlighted that the potential practical applications outlined above are currently speculative. Other than Experiment 6, all experiments were designed to address specific theoretical hypotheses and therefore sacrificed external validity in favor of greater experimental control. The form of deception studied here deviates from real-life deception in several important ways that limit the generalizability of the conclusions to a practical setting. Lying about a mock crime that one has merely watched or imagined performing might affect memory differently than when an individual has actually performed the event in question. It is a well-established finding that memory is superior for self-performed events compared to events that are only read or observed (Engelkamp, 1998; Mulligan & Hornstein, 2003). It is therefore possible that memory for self-performed actions will be more resistant to distortion following deception. Almost all existing research into the effects of lying on memory requests that participants merely observe a mock crime video. An important question for future research is therefore whether memory is affected by lying to the same extent when actions are performed, rather than merely witnessed or imagined.

Perhaps even more important, future research should consider how social aspects of deception might moderate its effects on memory. The experiments within this thesis were conducted from a cognitive perspective and therefore focused on the memorial implications of the cognitive processes recruited in fabrication. However, deception is essentially a social phenomenon. The stakes of being exposed as a liar in criminal settings are very high and such circumstances therefore inspire great motivation to remain undetected. Of course, ethical restrictions prevent manipulation of the stakes of lying in any way that approximates a real life criminal situation and the inability to determine the ground truth in real criminal cases limit the conclusions that can be drawn from the field. We do, however, know that motivation to be believed can increase the likelihood of deception-induced memory distortion. This has been studied in the context of romantic attraction: People are typically more motivated to impress people that they deem attractive. In one study, participants more often misremembered agreement ratings they had previously given in response to certain statements to align more closely with the ratings of an attractive other, but not an unattractive other (Brady & Lord, 2013). Thus, the motivation to appear more similar to an attractive other led people to misremember that their prior opinions were actually discordant with that individual’s opinions. Social aspects of deception like the stakes and motivation will therefore likely moderate the effect of lying on memory.

6.3. Concluding remarks

Memory is extraordinarily important; in many cases, it is the only record we have of the past. And because of that, we have the power to shape other people's beliefs about the past through the information we share: We can omit, exaggerate, embellish or outright fabricate information to manipulate others' beliefs. But perhaps we do not realize that in shaping others' beliefs, we can unwittingly alter our own and lose an accurate representation of what we previously knew. A faithful representation of the past is more important in some contexts than others. The criminal justice system is perhaps the most extreme example of where the truth matters, where even tiny seemingly insignificant details can prove critical. It is also in this context where lying is perhaps most prevalent. Thus, not only are there more opportunities for lying to alter memory, but the consequences of this may also be the most severe. The experiments contained within this thesis show that, under some circumstances, lying can impair memory for the truth and suggest that this may indeed be problematic for the criminal justice system. The primary focus was developing theory for how deception-induced memory distortion might occur. Ultimately, a good theory can guide future empirical research to determine when lying might be problematic for memory and can inform the design of interventions to mitigate that. While there is still much to explore and understand in this area of research, this thesis advances our theoretical understanding of how we might inadvertently be the source of our own memory distortion through the act of deception.

References

Abe, N. (2011). How the brain shapes deception: an integrated review of the literature. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *17*(5), 560–574.

Ackil, J. K., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology*, *34*(6), 1358–1372.

Ackil, J. K., & Zaragoza, M. S. (2011). Forced fabrication versus interviewer suggestions: Differences in false memory depend on how memory is assessed. *Applied Cognitive Psychology*, *25*(6), 933–942.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(5), 1063–1087.

Anderson, S. F., & Kelley, K. (2019). *BUCCS: Bias and Uncertainty Corrected Sample Size*. R Package. Version 1.1.0, https://CRAN.R-project.org/package=BUCSS.

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-Size Planning for More Accurate Statistical Power: A Method Adjusting Sample Effect Sizes for Publication Bias and Uncertainty. *Psychological Science*, *28*(11), 1547–1562.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. New York: Cambridge University Press.

Ben-Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in skin conductance. *Psychophysiology*, *22*(3), 292–299.

Ben-Shakhar, G., & Elaad, E. (2002). Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: a reexamination. *The Journal of Applied Psychology*, *87*(5), 972–977.

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *The Journal of Applied Psychology*, *88*(1), 131–151.

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology*, *94*(1), 1–11.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: a meta-analytic review. *Memory & Cognition*, *35*(2), 201–210.

Besken, M. (2018). Generating lies produces lower memory predictions and higher memory performance than telling the truth: Evidence for a metacognitive illusion. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(3), 465–484.

Bickel, P. J., Hammel, E. A., & O'Connell, J. W. (1975). Sex bias in graduate admissions: data from Berkeley. *Science, 187*(4175), 398–404.

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, *3*(2), 77–88.

Brady, S. E., & Lord, C. G. (2013). When Liars Fool Themselves: Motive to Impress Alters Memory for One's Own Past Evaluative Actions. *Social Cognition*, *31*(5), 599–612.

Buchli, D. R., Storm, B. C., & Bjork, R. A. (2016). Explaining retrieval-induced forgetting: A change in mental context between the study and restudy practice phases is not sufficient to cause forgetting. *Quarterly Journal of Experimental Psychology*, *69*(6), 1197–1209.

Camp, G., & Wesstein, H. (2012). Can questioning induce forgetting? Retrieval-induced forgetting of eyewitness information. *Applied Cognitive Psychology*, *26*(3), 431–435.

Chan, J. C. K., Erdman, M. R., & Davis, S. D. (2015). Retrieval induces forgetting, but only when nontested items compete for retrieval: Implication for interference, inhibition, and context reinstatement. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*(5), 1298.

Christianson, S. Å., & Bylin, S. (1999). Does simulating amnesia mediate genuine forgetting for a crime event? *Applied Cognitive Psychology*, *13*, 495–511.

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex*, *19*(7), 1557–1566.

Chrobak, Q. M., & Zaragoza, M. S. (2008). Inventing stories: forcing witnesses to fabricate entire fictitious events leads to freely reported false memories. *Psychonomic Bulletin & Review*, *15*(6), 1190–1195.

Chrobak, Q. M., & Zaragoza, M. S. (2013). When forced fabrications become truth: causal explanations and false memory development. *Journal of Experimental Psychology: General*, *142*(3), 827–844.

Chrobak, Q., & Zaragoza, M. S. (2009). The cognitive consequences of forced fabrication: Evidence from studies of eyewitness suggestibility. In W. Hirstein (Ed.), *Confabulation: Views from neuroscience, psychiatry, psychology and philosophy* (pp. 67–90). Oxford: Oxford University Press.

Cima, M., Nijman, H., Merckelbach, H., Kremer, K., & Hollnack, S. (2004). Claims of crime-related amnesia in forensic patients. *International Journal of Law and Psychiatry*, *27*(3), 215–221.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology. General*, *125*(2), 159–180.

Curran, T., & Doyle, J. (2011). Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. *Journal of Cognitive Neuroscience*, *23*(5), 1247–1262.

Debey, E., De Houwer, J., & Verschuere, B. (2014). Lying relies on the truth. *Cognition*, *132*(3), 324–334.

Debey, E., Ridderinkhof, R. K., De Houwer, J., De Schryver, M., & Verschuere, B. (2015). Suppressing the truth as a mechanism of deception: Delta plots reveal the role of response inhibition in lying. *Consciousness and Cognition*, *37*, 148–159.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, *44*(3), 837–845.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118.

Dewhurst, S. A., & Hitch, G. J. (1999). Cognitive effort and recollective experience in recognition memory. *Memory, 7*(2), 129–146.

Dianiska, R. E., Cash, D. K., Lane, S. M., & Meissner, C. A. (2019). The Reciprocal Nature of Lying and Memory: Memory Confabulation and Diagnostic Cues to Deception. In T. Docan-Morgan (Ed.), *The Palgrave Handbook of Deceptive Communication* (pp. 347–365). Cham: Springer International Publishing.

Dodson, C. S., Koutstaal, W., & Schacter, D. L. (2000). Escape from illusion: reducing false memories. *Trends in Cognitive Sciences*, *4*(10), 391–397.

Duran, N. D., Dale, R., & McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychonomic Bulletin & Review*, *17*(4), 486–491.

Elaad, E. (2011). Validity of the Concealed Information Test in realistic contexts. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 171–186). New York: Cambridge University Press.

Engelkamp, J. (1998). *Memory for actions*. Hove: Psychology Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage Publications Ltd.

Finke, R. A., Johnson, M. K., & Shyi, G. C. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, *16*(2), 133–137.

Foley, M. A., & Foley, H. J. (2007). Source-monitoring judgments about anagrams and their solutions: evidence for the role of cognitive operations information in memory. *Memory & Cognition*, *35*(2), 211–221.

Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, Ben-Shakhar, Gershon, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 27–45). Cambridge: Cambridge University Press.

Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the Guilty Actions Test. *Biological Psychology*, *83*(2), 101–107.

Gawrylowicz, J., Fairlamb, S., Tantot, E., Qureshi, Z., Redha, A., & Ridley, A. M. (2016). Does Practice Make the Perfect Liar? The Effect of Rehearsal and Increased Cognitive Load on Cues to Deception. *Applied Cognitive Psychology*, *30*(2), 250–259.

Goff, L. M., & Roediger, H. L., 3rd. (1998). Imagination inflation for action events: repeated imaginings lead to illusory recollections. *Memory & Cognition*, *26*(1), 20–33.

Granhag, P. A., & Strömwall, L. A. (2000). Deception detection: Examining the consistency heuristic. *New Trends in Criminal Investigation and Evidence*, *2*, 309–321.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin*, *75*(6), 424–429.

Gronau, N., Elber, L., Satran, S., Breska, A., & Ben-Shakhar, G. (2015). Retroactive memory interference: a potential countermeasure technique against psychophysiological knowledge detection methods. *Biological Psychology*, *106*, 68–78.

Hart, R. E., & Schooler, J. W. (2006). Increasing belief in the experience of an invasive procedure that never happened: the role of plausibility and schematicity. *Applied Cognitive Psychology*, *20*(5), 661–669.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior*, *29*(4), 469–484.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *27*(1), 46–51.

Haydock, L. (2018, October 15). Witness intimidation revealed. Retrieved October 25, 2019 from https://www.facebook.com/notes/dispatches-channel-4/witness-intimidation-revealed/2738298862862643/.

Hekkanen, S. T., & McEvoy, C. (2002). False memories and source-monitoring problems: criterion differences. *Applied Cognitive Psychology*, *16*(1), 73–85.

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(5), 1195–1209.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. Wiley New York.

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing Unwanted Autobiographical Memories Reduces Their Automatic Influences: Evidence From Electrophysiology and an Implicit Autobiographical Memory Test. *Psychological Science, 26*(7), 1098–1106.

Hu, X., Chen, H., & Fu, G. (2012). A repeated lie becomes a truth? The effect of intentional control and training on deception. *Frontiers in Psychology*, *3*, 488.

Informing justice: The disturbing use of jailhouse informants. (n.d.). Retrieved from https://www.innocenceproject.org/informing-injustice/.

Ioannou, M., & Hammond, L. (2015). The Detection of Deception Within Investigative Contexts: Key Challenges and Core Issues. *Journal of Investigative Psychology and Offender Profiling*, *12*(2), 107–118.

Iversen, S. D., & Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Experimental Brain Research*, *11*(4), 376–386.

JASP Team (2019). JASP (Version 0.9.1.0) [Computer software].

Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. The Journal of Problem Solving, *7*(1), 2.

Jian, Z., Zhang, W., Tian, L., Fan, W., & Zhong, Y. (2019). Self-Deception Reduces Cognitive Load: The Role of Involuntary Conscious Memory Impairment. *Frontiers in Psychology, 10*, 60.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology. General*, *117*(4), 371–376.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.

Johnson, R., Jr, Barnhardt, J., & Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia*, *42*(7), 878–901.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, *42*(3), 419–427.

Klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2016). Orienting versus inhibition in the Concealed Information Test: Different cognitive processes drive different physiological measures. *Psychophysiology*, *53*(4), 579–590.

Klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2017). Unraveling the roles of orienting and inhibition in the Concealed Information Test. *Psychophysiology*, *54*(4), 628–639.

Knott, L. M., & Dewhurst, S. A. (2007). The effects of divided attention at study and test on false recognition: a comparison of DRM and categorized lists. *Memory & Cognition*, *35*(8), 1954–1965.

Lancaster, G. L. J., Vrij, A., Hope, L., & Waller, B. (2013). Sorting the Liars from the Truth Tellers: The Benefits of Asking Unanticipated Questions on Lie Detection. *Applied Cognitive Psychology*, *27*(1), 107–114.

Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology*, *20*(2), 199–212.

Leins, D. A., Fisher, R. P., & Ross, S. J. (2013). Exploring liars' strategies for creating deceptive reports. *Legal and Criminological Psychology*, *18*(1), 141–151.

Leins, D., Fisher, R. P., Vrij, A., Leal, S., & Mann, S. (2011). Using sketch drawing to induce inconsistency in liars. *Legal and Criminological Psychology*, *16*(2), 253–265.

Levine, T. R., Pete Blair, J., & Carpenter, C. J. (2018). A critical look at meta-analytic evidence for the cognitive approach to lie detection: A re-examination of Vrij, Fisher, and Blank (2017). *Legal and Criminological Psychology*, Vol. 23, pp. 7–19.

Lindsay, D. S. (2008). Source Monitoring. In J. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (Vol. 2, pp. 325–348). Oxford: Elsevier.

Lindsay, D. S. (2014). Memory source monitoring applied. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 59–75). Sage London, England.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*(3), 349–358.

Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366.

Luke, T. J. (2019). Lessons From Pinocchio: Cues to Deception May Be Highly Exaggerated. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *14*(4), 646–671.

Lykken, D. T. (1974). Psychology and the lie detector industry. *The American Psychologist*, *29*(10), 725–739.

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. New York: McGraw Hill.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral & Social Sciences*, *1*(3), 86–92.

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: forgetting as a consequence of remembering. *Applied Cognitive Psychology*, *16*(2), 135–149.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199.

Maisto, A. A., & Queen, D. E. (1992). Memory for pictorial information and the picture superiority effect. *Educational Gerontology*, *18*(2), 213–223.

Mangiulli, I., Van Oorsouw, K., Curci, A., & Jelicic, M. (2019). Retrieval-Induced Forgetting in the Feigning Amnesia for a Crime Paradigm. *Frontiers in Psychology*, *10*, 928.

Mangiulli, I., Van Oorsouw, K., Curci, A., Merckelbach, H., & Jelicic, M. (2018). Feigning Amnesia Moderately Impairs Memory for a Mock Crime Video. *Frontiers in Psychology*, *9*, 625.

Marion, S. B., & Burke, T. M. (2013). False alibi corroboration: witnesses lie for suspects who seem innocent, whether they like them or not. *Law and Human Behavior*, *37*(2), 136–143.

Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 26, pp. 1124–1140.

McBride, D. M., & Dosher, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: a process dissociation analysis. *Consciousness and Cognition*, *11*(3), 423–460.

McDonough, I. M., & Gallo, D. A. (2008). Autobiographical elaboration reduces memory distortion: cognitive operations and the distinctiveness heuristic. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(6), 1430–1445.

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51*(9), 879–904.

Migueles, M., & García-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology*, *21*(9), 1157–1172.

Mulligan, N. W., & Hornstein, S. L. (2003). Memory for actions: self-performed tasks and the reenactment effect. *Memory & Cognition*, *31*(3), 412–421.

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: a meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *140*(5), 1383–1409.

Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: the role of memory for crime details. *Psychophysiology*, *48*(6), 733–744.

Nieznański, M. (2011). Generation difficulty and memory for source. *Quarterly Journal of Experimental Psychology*, *64*(8), 1593–1608.

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review*, *114*(4), 887–953.

Orwell, G. (1954). *1984*. London, England: Penguin Books.

Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 253–275). Cambridge: Cambridge University Press.

Otgaar, H., & Baker, A. (2018). When lying changes memory for the truth. *Memory*, *26*(1), 2–14.

Otgaar, H., Howe, M. L., Memon, A., & Wang, J. (2014). The development of differential mnemonic effects of false denials and forced confabulations. *Behavioral Sciences & the Law*, *32*(6), 718–731.

Otgaar, H., Howe, M. L., Smeets, T., & Wang, J. (2016). Denial-Induced Forgetting: False Denials Undermine Memory, But External Denials Undermine Belief. *Journal of Applied Research in Memory and Cognition*, *5*(2), 168–175.

Otgaar, H., Romeo, T., Ramakers, N., & Howe, M. L. (2017). Forgetting having denied: The “amnesic” consequences of denial. *Memory & Cognition*, *46*, 520-529.

Paas, F. G., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*, 419–430.

Paige, L. E., Fields, E. C., & Gutchess, A. (2018). Influence of age on the effects of lying on memory. *Brain and Cognition*, *133*, 42-53.

Payne, D. G., Neely, J. H., & Burns, D. J. (1986). The generation effect: further tests of the lexical activation hypothesis. *Memory & Cognition*, *14*(3), 246–252.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

Pérez-Mata, M. N., Read, J. D., & Diges, M. (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory*, *10*(3), 161–177.

Perfect, T. J., Moulin, C. J. A., Conway, M. A., & Perry, E. (2002). Assessing the inhibitory account of retrieval-induced forgetting with implicit-memory tests. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(6), 1111–1119.

Peters, M. J. V., Jelicic, M., Gorski, B., Sijstermans, K., Giesbrecht, T., & Merckelbach, H. (2008). The corrective effects of warning on false memories in the DRM paradigm are limited to full attention conditions. *Acta Psychologica*, *129*(2), 308–314.

Pezdek, K., Blandon-Gitlin, I., Lam, S., Hart, R. E., & Schooler, J. W. (2006). Is knowing believing? The role of event plausibility and background knowledge in planting false beliefs about the personal past. *Memory & Cognition*, Vol. 34, pp. 1628–1635.

Pezdek, K., Lam, S. T., & Sperry, K. (2009). Forced confabulation more strongly influences event memory if suggestions are other-generated than self-generated. *Legal and Criminological Psychology*, *14*(2), 241–252.

Pezdek, K., Sperry, K., & Owens, S. M. (2007). Interviewing witnesses: The effect of forced confabulation on event memory. *Law and Human Behavior, 31*(5), 463–478.

Pickel, K. L. (2004). When a lie becomes the truth: the effects of self-generated misinformation on eyewitness memory. *Memory*, *12*(1), 14–26.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2019). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-143, https://CRAN.R-project.org/package=nlme.

Polage, D. C. (2004). Fabrication deflation? The mixed effects of lying on memory. *Applied Cognitive Psychology*, *18*(4), 455–465.

Polage, D. C. (2012). Fabrication inflation increases as source monitoring ability decreases. *Acta Psychologica*, *139*(2), 335–342.

Polage, D. C. (2018). Liar, liar: Consistent lying decreases belief in the truth. *Applied Cognitive Psychology*, *20*, 1063.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, *1*(1), 125–126.

Qualtrics. (2005). Provo, Utah, USA. Available at: https://www.qualtrics.com.

Raaijmakers, J. G. W., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language*, *68*(2), 98–122.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riefer, D. M., Chien, Y., & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *Quarterly Journal of Experimental Psychology*, *60*(10), 1389–1405.

Rindal, E. J. (2017). *The Effect of Repeated Lying on False Memory Development* (PhD, Kent State University, Kent, USA). Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=kent1500577222767892.

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.

Romeo, T., Otgaar, H., Smeets, T., Landstrom, S., & Boerboom, D. (2018). The impact of lying about a traumatic virtual reality experience on memory. *Memory & Cognition*, *47*(3), 485-495.

Rowland, C. A., Bates, L. E., & DeLosh, E. L. (2014). On the reliability of retrieval-induced forgetting. *Frontiers in Psychology, 5*, 1–10.

Schweitzer, M. E., Brodt, S. E., & Croson, R. T. A. (2002). Seeing and believing: Visual access and the strategic use of deception. *International Journal of Conflict Management*, *13*(3), 258–375.

Seymour, T. L., & Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Applied Psychophysiology and Biofeedback*, *34*(3), 177–187.

Sharman, S. J., & Scoboria, A. (2009). Imagination equally influences false memories of high and low plausibility events. *Applied Cognitive Psychology*, *23*(6), 813–827.

Shaw, D. J., Vrij, A., Leal, S., Mann, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2013). Expect the Unexpected? Variations in Question Type Elicit Cues to Deception in Joint Interviewer Contexts. *Applied Cognitive Psychology*, *27*(3), 336–343.

Shaw J. S., Bjork R. A., Handal A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review*, *2*, 249–253.

Sheridan, M. R., & Flowers, K. A. (2010). Reaction times and deception - the lying constant. *International Journal of Psychological Studies*, *2*(2), 41–51.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592.

Smeets, T., Candel, I., & Merckelbach, H. (2004). Accuracy, completeness, and consistency of emotional memories. *The American Journal of Psychology*, *117*(4), 595–609.

Spence, S. A., Farrow, T., Herford, A., Zheng, Y., Wilkinson, I. D., Brook, M. L., & Woodruff, P. W. R. (2001). A preliminary description of the behavioural and functional anatomical correlates of lying. *NeuroImage*, *13*(6), 477.

Sporer, S. L. (2016). Deception and Cognitive Load: Expanding Our Horizon with a Working Memory Model. *Frontiers in Psychology*, *7*, 420.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods*, *Instruments, & Computers*, *31*(1), 137–149.

Storm, B. C. (2011). Retrieval-induced forgetting and the resolution of competition. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 89–105). New York: Psychology Press.

Storm, B. C., Angello, G., Buchli, D. R., Koppel, R. H., Little, J. L., & Nestojko, J. F. (2015). A Review of Retrieval-Induced Forgetting in the Contexts of Learning, Eyewitness Memory, Social Cognition, Autobiographical Memory, and Creative Cognition. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 62, pp. 141–194). Academic Press.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2007). When intended remembering leads to unintended forgetting. *Quarterly Journal of Experimental Psychology*, *60*(7), 909–915.

Suchotzki, K., Crombez, G., Smulders, F. T. Y., Meijer, E., & Verschuere, B. (2015). The cognitive mechanisms underlying deception: an event-related potential study. *International Journal of Psychophysiology*, *95*(3), 395–405.

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428–453.

Sugimori, E., & Tanno, Y. (2010). The effects of cognitive activity and perceptual details on speech source monitoring. *British Journal of Psychology*, *101*(4), 777–790.

Sun, S.-Y., Mai, X., Liu, C., Liu, J.-Y., & Luo, Y.-J. (2011). The processes leading to deception: ERP spatiotemporal principal component analysis and source analysis. *Social Neuroscience*, *6*(4), 348–359.

Sun, X., Punjabi, P. V., Greenberg, L. T., & Seamon, J. G. (2009). Does feigning amnesia impair subsequent recall? *Memory & Cognition*, *37*(1), 81–89.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296.

Taconnat, L., & Isingrini, M. (2004). Cognitive operations in the generation effect on a recall test: role of aging and divided attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(4), 827–837.

Takarangi, M. K. T., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: the development of a new stimulus set. *Applied Cognitive Psychology*, *20*(5), 583–590.

Time to tackle witness intimidation. (2003, January 11). Retrieved from http://news.bbc.co.uk/1/hi/uk/2645867.stm.

The Global Deception Research Team. (2006). A World of Lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74.

Troyer, A. K., & Craik, F. I. (2000). The effect of divided attention on memory for items and their context. *Canadian Journal of Experimental Psychology*, *54*(3), 161–171.

Tversky, B., & Marsh, E. J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, *40*(1), 1–38.

Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., & Spruyt, A. (2012). Learning to lie: Effects of practice on the cognitive cost of lying. *Frontiers in Psychology*, *3*, 1–8.

Van Damme, I., Menten, J., & d'Ydewalle, G. (2010). The effect of articulatory suppression on implicit and explicit false memory in the DRM paradigm. *Memory*, *18*(8), 822–830.

Van Duisen, M. (2014, August 17). 10 crimes solved by a tiny piece of evidence. Retrieved December 15, 2019, from https://listverse.com/2014/08/17/10-crimes-solved-by-a-tiny-piece-of-evidence/.

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., … al., E. (2019). *The JASP Guidelines for Conducting and Reporting a Bayesian Analysis*. https://doi.org/10.31234/osf.io/yqxfr

Van Oorsouw, K., & Merckelbach, H. (2004). Feigning amnesia undermines memory for a mock crime. *Applied Cognitive Psychology*, *18*(5), 505–518.

Van Oorsouw, K., & Merckelbach, H. (2006). Simulating amnesia and memories of a mock crime. *Psychology, Crime & Law: PC & L*, *12*(3), 261–271.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language*, *50*(3), 289–335.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage*, *14*(6), 1302–1308.

Veling, H., & van Knippenberg, A. (2004). [Review of *Remembering can cause inhibition: retrieval-induced inhibition as cue independent process*]. *Journal of experimental psychology. Learning, memory, and cognition*, *30*(2), 315–318.

Venables, P. H., & Mitchell, D. A. (1996). The effects of age, sex and time of testing on skin conductance activity. *Biological Psychology*, *43*(2), 87–101.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). *Memory detection: Theory and application of the concealed information test*. Cambridge University Press.

Verschuere, B., Köbis, N. C., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2018). Taxing the Brain to Uncover Lying? Meta-analyzing the Effect of Imposing Cognitive Load on the Reaction-Time Costs of Lying. *Journal of Applied Research in Memory and Cognition, 7*(3), 462-469.

Vieira, K. M., & Lane, S. M. (2013). How you lie affects what you remember. *Journal of Applied Research in Memory and Cognition*, *2*(3), 173–178.

Visu-Petra, G., Jurje, O., Opre, A., Miclea, M., & Visu-Petra, L. (2014). Which came first, the truth or the lie? Effects of order and habituation in the reaction time-based concealed information test. *Cognition, Brain, Behavior, 18*(4), 283–297.

Volz, K., Bahr, L.-M., Heinrichs, M., Vaitl, D., & Ambach, W. (2018). The Concealed Information Test is Susceptible to Misleading Information. *Journal of Forensic Sciences*, *63*(5), 1419–1428.

von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *The Behavioral and Brain Sciences*, *34*(1), 1–16.

Vredeveldt, A., van Koppen, P. J., & Granhag, P. A. (2014). The Inconsistent Suspect: A Systematic Review of Different Types of Consistency in Truth Tellers and Liars. In R. Bull (Ed.), *Investigative Interviewing* (pp. 183–207). New York, NY: Springer New York.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. West Sussex: John Wiley & Sons.

Vrij, A., Blank, H., & Fisher, R. P. (2018). A re-analysis that supports our main results: A reply to Levine et al. *Legal and Criminological Psychology*, *23*(1), 20–23.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, *5*(1-2), 39–43.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1), 1–21.

Vrij, A., Granhag, P. A., & Mann, S. (2010). Good Liars. *The Journal of Psychiatry & Law*, *38*(1-2), 77–98.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the Liars: Toward a Cognitive Lie Detection Approach. *Current Directions in Psychological Science*, *20*(1), 28–32.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253–265.

Vrij, A., Mann, S., Leal, S., & Fisher, R. (2010). “Look into my eyes”: Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law*, *16*(4), 327–348.

Wade, K. A., & Garry, M. (2005). Strategies for verifying false autobiographical memories. *The American Journal of Psychology*, *118*(4), 587–602.

Wade, K. A., Rowthorn, H., & Sukumar, D. (2017). The Malleability of Memory. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference (second edition)* (pp. 553–570). Elsevier.

Walczyk, J. J., Griffith, D. A., Yates, R., Visconte, S. R., Simoneaux, B., & Harris, L. L. (2012). Lie Detection by Inducing Cognitive Load: Eye Movements and Other Cues to the False Answers of “Witnesses” to Crimes. *Criminal Justice and Behavior*, *39*(7), 887–909.

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, *34*, 22–36.

Walczyk, J. J., Igou, F. P., Dixon, A. P., & Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Frontiers in Psychology*, *4*, 14.

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive Lie Detection: Response Time and Consistency of Answers as Cues to Deception. *Journal of Business and Psychology*, *24*(1), 33–49.

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Applied Cognitive Psychology*, *17*(7), 755–774.

Walczyk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M. I. N., & Zha, P. (2005). Lying person-to-person about life events: A cognitive framework for lie detection. *Personnel Psychology*, *58*(1), 141–170.

Williams, E. J., Bott, L. A., Patrick, J., & Lewis, M. B. (2013). Telling lies: the irrepressible truth? *PloS One*, *8*(4), e60713.

Wright, P. (2017, January 10). Oliver Tetlow killing: London rapper Cbiz cleared of murder after trial collapses. Retrieved from https://www.ibtimes.co.uk/oliver-tetlow-killing-london-rapper-cbiz-cleared-murder-after-trial-collapses-1600270.

Wright, D. B., Loftus, E. F., & Hall, M. (2001). Now you see it; now you don't: inhibiting recall and recognition of scenes. *Applied Cognitive Psychology*, *15*(5), 471–482.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517.

Zaragoza, M., Payment, K. E., Ackil, J. K., Drivdahl, S. B., & Beck, M. (2001). Interviewing witnesses: forced confabulation and confirmatory feedback increase false memories. *Psychological Science*, *12*(6), 473–477.

Zaragoza, M., Rich, P., Rindal, E., & DeFranco, R. (2017). Forced fabrication and false eyewitness memories. In R. A. Nash & J. Ost (Eds.), *False and Distorted Memories* (pp. 82–96). Oxon: Routledge.

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

Appendices

Appendix 1.

Experiment 1a required subjects to lie by fabricating an alternative category member beginning with the same letter as the item learned. We therefore conducted a pilot questionnaire to ensure that participants were able to generate appropriate alternatives for each item. The main experiment required 6 categories each composed of 6 items. We piloted 8 categories, each containing 8 items.

Participants (n=18) were shown each category-item pair and asked to generate at least one alternative category member beginning with the same letter. Participants also rated the difficulty of generating an alternative for each pair on a scale of 0-100, where 0 represented ‘very easy’ and 100 represented ‘very difficult’.

Table A1 shows the percentage of participants who provided at least one alternative category member for each item. The 6 categories with the highest percentages were selected for use in the main experiment. For each category, the 6 items with the lowest difficulty ratings for generating alternatives were selected for the main experiment.

Table A1.

The mean percentage of participants who generated at least one alternative category member beginning with the same letter as the item provided, and difficulty ratings (scale 0-100) for the top 6 items in each category. Starred categories were used in the main experiment.

Category	Mean percentage of participants who generated correct alternatives (SD)	Mean difficulty rating (SD)
Alcohol	80 (18.97)	37.65 (11.56)
Body*	100 (0)	6.33 (3.56)
Clothing*	95 (6.49)	20.72 (15.85)
Country*	100 (0)	16.32 (5.51)
Furniture	86 (16.37)	42.67 (24.69)
Occupation*	99 (2.27)	17.4 (6.59)
Sport*	94 (6.85)	35.64 (11.14)
Vegetable*	93 (8.36)	30.85 (13.34)

Appendix 2.

Category-item pairs used in Experiment 1a. (C) indicates critical categories and (F) indicates filler categories.

Category							
Body (C)	Clothing (C)	Country (C)	Occupation (C)	Sport (C)	Vegetable (C)	Tool (F)	Color (F)
Brain	Boxers	Brazil	Cook	Tennis	Pepper	Hammer	Green
Eyes	Coat	Canada	Banker	Swimming	Lettuce	Ruler	Purple
Foot	Hat	Germany	Athlete	Hockey	Cabbage	Wrench	Blue
Head	Jacket	Italy	Dentist	Golf	Asparagus		
Lips	Shorts	Mexico	Professor	Bowling	Beetroot		
Shoulder	Tie	Spain	Secretary	Running	Spinach		

Appendix 3.

Items that participants were instructed to remember in Experiment 2 together with their associated actions for the crime-relevant and crime-irrelevant image sequences. “(filler)” refers to non-critical items for which memory was not tested.

Category	Item	Action
Crime-relevant	Keycard (filler)	Opens door to the office using the keycard.
	Wall	Blueprints for the device are projected on the wall.
	Phone camera	Takes a photo of the blueprints with a phone.
	Red mouse	Identifies computer to hack as the one with the red mouse
	USB drive	Inserts the USB drive into the computer.
	Password ‘star2016’	The IT system is accessed with this password.
	Security image of brain	Circles back of the brain to access the IT system.
	Key F8	The key pressed to copy the relevant files over to the USB stick.
	Hard drive	A hard drive is stolen from the desk.
	Programming code (filler)	Programming code is printed and taken.
Crime-irrelevant	Rubber gloves (filler)	Puts on rubber gloves.
	Toilet brush	Scrubs toilet with a brush.
	Dustpan and brush	Sweeps floor with dustpan and brush.
	Mop	Cleans floor with mop.
	Caution sign	Puts out sign.
	Black bin	Empties the black bin.
	Window cleaner (Mr. Muscle)	Cleans the window using Mr. Muscle cleaner.
	Water dispenser	Wipes the water dispenser.
	Feather duster	Clears cobwebs with the feather duster.

Vacuum cleaner (filler)	Vacuums the carpet.

Appendix 4.

When developing the mock crime stimuli, three video sequences were filmed: 2 crime-relevant and one crime-irrelevant. This was to ensure that there was a choice in stimuli to match baseline memory for the crime-relevant and crime-irrelevant sequences as closely as possible. Ten still images were selected from each video as described in Section 2.11.2. Each image included an item to be remembered (example stimuli are shown in Figure 2.2). The crime-relevant image sequences depicted either a workshop break in to steal a device or an office break in to steal important documents. The crime-irrelevant sequence depicted a cleaning scenario. A pilot experiment was conducted to test participants' baseline memory for the image sequences and to determine which crime-relevant sequence was most appropriate for use in the main experiment.

In the pilot experiment, participants (n=51) viewed the three image sequences and were instructed to memorize the items highlighted in red. The order in which the image sequences were shown was counterbalanced across participants. After a 10-minute distractor phase, participants completed a cued-recall test in which they were presented with a question addressing the 30 critical items from the 3 image sequences. The questions asked were the same as those to be used in the retrieval practice phase of the main experiment. Questions were presented in a random order.

A one-way repeated-measures ANOVA indicated the proportion of items correctly recalled differed between the image sequences ($F(2,96) = 6.26$, $p = .003$). Bonferroni-adjusted pairwise comparisons revealed that participants recalled significantly more items from the crime-irrelevant cleaning sequence ($M = .74$, $SD = .18$) than the workshop break-in sequence ($M = .64$, $SD = .20$; $p = .002$, 95%CI [.03, .17]). There was no significant difference in recall performance for the cleaning sequence and the office break-in ($M = .68$, $SD = .20$; $p = .10$, 95% CI [-.09, .14]. The office break-in was therefore selected as the crime-relevant image sequence for use in the main experiment.

Appendix 5.

Critical items used in Experiment 3 and their associated easy and hard questions.

Category	Item	Easy question	Hard question
Body part	Arm	What body part might you cross in front of your chest?	What body part contains the Axillary artery?
	Eye	What body part do you see with?	What body part does 'Blephartis' affect?
	Foot	What body part makes contact with the ground when you walk?	What body part does 'plantar fasciitis' affect?
	Head	What body part is cut off when someone is decapitated?	What body part has the Latin name 'Caput'?
	Leg	What body part is most likely to contain varicose veins?	What body part does peripheral arterial disease mainly affect?
	Mouth	What body part do you open for the dentist?	What body part contains the uvula?
	Nose	What body part do you smell with?	What body part was the subject of an opera written by Shostakovich?
	Toe	What body part might you stub?	What body part enables you to balance?
Country	USA	What country is Donald Trump president of?	What country would you find the ’Garden of the Gods’ in?
	Canada	What country is Vancouver in?	What country has the longest coastline in the world?
	France	What country was Napoleon from?	What country is also known as ‘the hexagon’?
	Japan	What country does sushi come from?	What country consists of more than 6800 islands?
	England	What country is Stonehenge in?	What country was the inventor of the fire extinguisher from?
	Germany	What country has Berlin as its capital?	What country has changed its capital city more than 7 times?

	Spain	What country is Barcelona in	What country was the mop invented in?
	Italy	What country is famous for its love of pizza and pasta?	What country has the lily as its national flower?
Fruit	Apple	What fruit is cider made from?	What fruit contains the poison cyanide in its seeds?
	Orange	What fruit makes the juice that is commonly abbreviated to ‘OJ’?	What fruit originated in 4000BC in Southeast Asia?
	Banana	What fruit is long and curved with a yellow skin?	What fruit is also known as the ‘fruit of the wise men’?
	Grape	What fruit is wine made from?	What fruit has over 8000 varieties?
	Strawberry	What fruit is typically eaten with cream in the Summer?	What fruit did Othello decorate Desdemona’s handkerchief with in the Shakespearean play?
	Pear	What fruit completes the phrase ‘a partridge in a ___ tree’?	What fruit did Ancient Greeks use to treat nausea?
	Kiwi	What fruit with green flesh is also the name of a flightless bird?	What fruit is also known as the ‘Chinese gooseberry’?
	Watermelon	What fruit is large with pink flesh, black seeds and a thick green skin?	What fruit originated in the Kalahari Desert?
Musical Instrument	Flute	What musical instrument is held to the side of the face and played by blowing across a hole?	What musical instrument, still played today, dates to Paleolithic times?
	Trumpet	What musical instrument’s name comes from the word ‘trompe’?	What musical instrument was originally used for military and religious purposes?
	Violin	What musical instrument is also known as a ‘fiddle’?	What musical instrument does an Archetier help to make?
	Drum	What musical instrument do you hit with sticks?	What musical instrument would you play a ratamacue on?

	Saxophone	What musical instrument was created by Adolphe Sax?	What musical instrument was Cannonball Adderley famous for playing?
	Guitar	What musical instrument typically has 6 strings and can be acoustic or electric?	What musical instrument was preceded by the gittern and vihuela?
	Clarinet	What musical instrument is a woodwind instrument that is typically black in color?	What musical instrument is made by the company ’Chadesh’?
	Piano	What musical instrument has 88 black and white keys?	What musical instrument are the Steinway family famous for making?
Sport	Football	What sport involves kicking a ball into a goal?	What sport is also referred to as ‘the beautiful game’?
	Basketball	What sport attracts tall players and involves throwing a ball into a hoop?	What sport do the Memphis Grizzlies belong to?
	Tennis	What sport involves two players hitting a ball over a net with a racket?	What sport originally had an hourglass-shaped court?
	Hockey	What sport involves moving a ball or puck into a goal using a long stick?	What sport originates from the term ‘shepherd’s crook’?
	Golf	What sport uses the term ‘a hole in one’?	What sport holds a tournament called’ The Solheim Cup’?
	Lacrosse	What sport, played on a pitch, requires its players to use nets at the end of a long stick?	What sport originated as a tribal game played by Native Americans?

Appendix 6.

Interview questions, the truthful answers, and the corresponding lies provided to subjects in Experiment 4 during interview for subjects under in the low and high cognitive load conditions. (C) indicates critical items and (P) indicates practice items.

Interview Question	Truth Item	Lie Item (low load)	Lie item (high load)
What did you use to break into the workshop?	A crowbar (P)	A SCREWDRIVER	A SCWRIVRDEER
What was the pattern to disable the alarm?	A Z pattern (C)	A LIGHTNING BOLT	A LIEHTNGING BLTO
What did you use to cover the CCTV camera?	Shaving foam (C)	A CLOTH	A CTOHL
Where was the device kept?	In a cupboard (C)	IN A DRAWER	IN A DARWRE
Where was the code for the combination lock?	On the perpetrator's watch (C)	ON SOME PAPER	ON SOME PRPEA
What was the device stored in?	A hardcase (C)	A BAG	A GBA
What did you use to cut the padlock?	Bolt cutters (C)	A HACKSAW	A HSAKCAW
What does the device do?	Holographic projection (C)	THERMAL IMAGING	TRHEAML IGMAING

What kind of bag did you put the device in?	A sports bag (C)	A BACKPACK	A BCKACAPK
How did you get into the office?	A keycard (C)	A NUMBER CODE	A NBUEMR CEDO
Where were the device blueprints shown?	On the wall (C)	ON A SCREEN	ON A SERCNE
How did you know what computer to hack?	It had a red mouse (C)	IT HAD A BLUE MOUSE	IT HAD A BEUL MUSOE
What did you take the photos with?	A phone camera (C)	A DIGITAL CAMERA	A DGTIAIL CAEMAR
What did you copy the computer files onto?	A USB stick (C)	AN EXTERNAL HARDDRIVE	AN EXNEARTL HARIDRVDE
What was the password for the computer?	STAR2017 (C)	SUN2016	NSU2016
What was the security hologram?	A brain (C)	A HEART	A HRETA
The folder on the computer you copied–what was it called?	Prototype8 (C)	BLUEPRINTS	BLUREIPTNS
What document did you print?	Programming code (P)	COMPANY ACCOUNTS	CMPONAY ACUONCTS

Appendix 7.

Interview questions, their truthful answers and corresponding lie answers for participants in the low load and high load conditions in Experiment 5. (C) indicates critical items and (P) indicates practice items.

Interview Question	Truth item (easy)	Truth item (hard)	Lie item (easy)	Lie item (hard)
What was Eric wearing? (C)	JENAS	ENSJA	OVEARLLS	ERLVOLAS
What did Eric eat? (C)	APLPE	LPAEP	PECAH	CAPHE
What magazine did Eric read? (C)	TMIE	ITEM	NEWWSEEK	WENSEKEW
Where did Eric read the note from the homeowner? (C)	HALWLAY	AWLHYAL	KITHCEN	NCEHKTI
What tool did Eric use in the kitchen? (C)	SCREDWRIVER	RSDWCVEIRRE	PLEIRS	RPILSE
What did the picture that Eric looked at show? (C)	THE LENAING TWOER	ALNNEIG ROEWT	EFIFEL TWOER	FLIEFE ROEWT
What was the state of the bed in the first bedroom? (P)	AMDE	DMEA	n/a	n/a

What did Eric test in the lounge? (C)	A LGIHT FITITNG	A GTLHI TNFIGIT	A POEWR SOCEKT	A EWPRO KOSTCE
What did Eric play? (P)	n/a	n/a	A VIEDO	A OEDVI
What color cap did Eric try on in the second bedroom? (C)	BLCAK	AKLCB	MAORON	OAROMN
What was the name of Eric's company? (C)	JR's ELECRTICIANS	JR'S ITELCERCINAS	JA'S ELECRTICIANS	JA'S ITELCERCINAS
On what did Eric check the time? (C)	WLAL CLCOK	LALW KCOCL	WTACH	THAWC
What did Eric steal from the first bedroom? (C)	ERARINGS	GERARINS	NECLKACE	KELCENAC
What did Eric look through in the lounge? (C)	A POHTO ALUBM	A OHTPO LBMAU	A JORUNAL	A NUJORLA
What color was Eric's van? (C)	BULE	LEBU	SIVLER	VIESRL
What did Eric find the housekey under? (C)	THE FLOEWRPOT	THE LEFWTPOOR	THE DOROMAT	THE RODMOTA

What color was the mug that was next to the papers that Eric rummaged through? (C)	WHTIE	ETWIH	BEGIE	EEGIB
What did Eric take from the fridge? (C)	CKOE	OEKC	PESPI	IESPP
What did Eric steal from the bathroom? (C)	PLILS	LISLP	PERUFME	RUEFEPM
What did Eric steal from the second bedroom? (C)	A RNIG	A IGRN	MNOEY	OYEMN

Appendix 8. Pilot study results for Experiment 4.

To create the fabrications for each interview question in Experiment 4, alternative false answers were generated for each question and a pilot survey was conducted online to ensure that the alternatives were plausible lies. Participants (n=21) were told about the mock crime and given a summary of what happens in the video. They were then shown the interview questions to be used in the main experiment, together with the correct answer (i.e., the item from the mock crime video). For example, participants saw 'Q: What was used to cover the CCTV camera? A: Shaving foam'. Participants were then given two alternative false answers to the question and indicated the plausibility of each false answer on a 7-point likert scale ranging from 1 ("I would definitely disbelieve this answer") to 7 ("I would definitely believe this answer"). The false answer that participants rated as most plausible was selected as the fabrication for the main experiment.

The mean plausibility rating for the selected false answers was 5.49 (out of 7; *SD* = 0.30). A rating of 3 or less indicated that participants would not believe the answer and therefore that the lie was implausible (3 corresponded to "I might disbelieve this answer"). All false answers were rated above 3, indicating that all selected answers were plausible fabrications.

A second pilot study was conducted to ensure that the anagrams used in the high load condition were considered more cognitively demanding to solve than merely reading the answer. Participants (n=18) were asked to solve anagrams that required them to switch two underlined letters (as used by Foley & Foley, 2007). However, participants did not rate this task as more cognitively demanding than the low cognitive load condition, which merely required them to read the deceptive answer. In fact, participants in the high cognitive load condition provided lower difficulty ratings (M = 2.67, SD = 0.87) than participants in the low cognitive load condition (M = 3.33, SD = 2.0), though these ratings did not significantly differ ($t(16) = 0.92$, p = .37, 95%CI[-0.87, 2.21]).

A further pilot study (n=20) was therefore conducted that tested two different ways of increasing cognitive load. Either 3 scrambled letters were underlined with a 10 second deadline for solving the anagram, or 4+ letters were scrambled. Participants solved each anagram and then rated how much cognitive effort they required to solve the anagrams on a scale of 1-9. Participant in the 4+ letter condition rated the task as significantly harder than Participants in the 3 letter deadline condition (M_{4+} = 5.0, *SD* = 1.56; M_3 = 3.30, *SD* = 1.49; $t(18) = -2.49$, $p = .02$, 95%CI[-3.14, -0.26]). Anagrams in the main experiment therefore had 4+ letters underlined.

Appendix 9. Source memory analysis for Experiment 5.

Section 4.7.4 outlines the findings of the source analysis for the lies that participants reported during interview. The results of the analyses for the remaining item types are reported below (items lied about, items reported truthfully, and items not asked about).

A series of Poisson loglinear generalized estimating equations (GEE) were performed to examine the distribution of source responses for each type of item and determine the types of source errors participants made (depicted in Figure A9.1). Separate GEE models were performed on each type of item—control items, lies that participants reported during interview, items from the crime that participants lied about during interview, and items from the crime that participants reported truthfully during interview—with Cognitive Load and the source option selected as predictors. From this, the likelihood that participants selected the correct source option compared to the remaining 4 source options could be calculated. Table A9.1 shows the odds ratios (ORs) for the change in the count for each source option compared to the correct option. The findings for each type of item are summarized below (excluding the items reported during interview, as these results are reported in Section 4.7.4).

Control items. The top left panel of Figure A9.1 shows the distribution of source responses for items shown in the mock crime that participants were not asked about during interview (control items). Given that these items were not discussed during interview, the cognitive load manipulation was not expected to affect source memory for these items. As expected, the distributions for the source responses were similar across all cognitive load conditions and participants selected the correct source option significantly more often than all other options, regardless of cognitive load ($\chi^2(4) = 169.59$, $p < .001$; see Table A9.1 for ORs). The GEE confirmed that there was no interaction between Cognitive Load and the source option selected ($\chi^2(8) = 4.20$, $p = .84$). This provides evidence that the differences in source memory performance between cognitive load conditions seen for other item types are indeed attributable to the cognitive load manipulation during interview.

Items lied about. The bottom left panel of Figure A9.1 shows the distribution of source responses for items that participants lied about during interview. We can see that participants selected the correct source option more often than all other options, regardless of cognitive load ($\chi^2(4) = 154.71$, $p < .001$; see Table A9.1 for ORs). The type of cognitive load imposed during interview did not affect source memory for items lied about. The GEE confirmed that there was no interaction between Cognitive Load and the source option selected ($\chi^2(8) = 6.13$, $p = .63$). Thus, participants typically did not forget that these items originated from the crime after lying about them and remembered that they lied about them during interview, regardless of cognitive load.

Items reported truthfully. We can see from the bottom right panel of Figure A9.1 that the distribution of source responses for items reported truthfully during interview differs depending on the cognitive load imposed during interview: the GEE revealed a significant interaction between Cognitive Load and the source option selected ($\chi^2(8) = 35.18$, $p < .001$). To break down this interaction, separate GEE models were performed for each cognitive load condition. The differences in the option selected are most pronounced for participants interviewed under high intrinsic load. These participants selected the correct option significantly more often than all other responses and showed the best source performance (see Table A9.1 for ORs).

Participants interviewed under low load also selected the correct option significantly more often than all other responses, but the effect size for the difference in the counts was smaller than participants interviewed under high intrinsic load, as indicated by smaller ORs. Participants interviewed under high extraneous load also rarely forgot that these items came from the crime, but were no less likely to select option 2 ("This item was shown in the mock crime, but I lied when I was asked about it at interview") compared to the correct answer, indicating that participants interviewed under high extraneous load sometimes misremembered how they spoke about the items during interview. Overall, regardless of the cognitive load imposed during interview, participants rarely forgot that items originated from the crime when they reported them truthfully during interview.

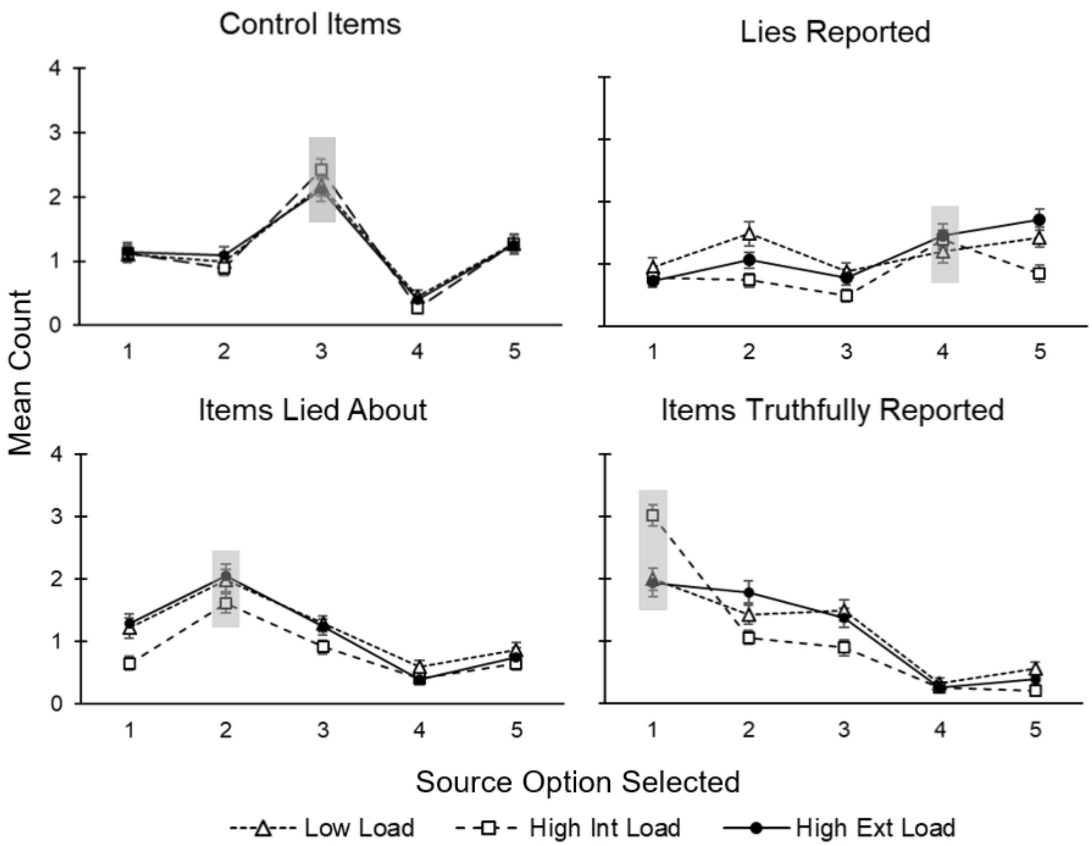


Figure A9.1. The distribution of source attributions for items that participants were not interviewed about (top left panel), lies reported during interview (top right panel), items lied about during interview (bottom left panel) and items reported truthfully during interview (bottom right panel). The correct source option for each type of item is highlighted in grey. Attributions of ≤3 represent the judgement that the item was shown in the mock crime and attributions of >3 represent the judgement that the item was not shown in the mock crime: 1 = "This was shown in the mock crime and I gave it as a truthful answer to a question at interview", 2 = "This was shown in the mock crime, but I lied when I was asked about it at interview", 3 = "This was shown in the mock crime, but I was NOT asked a question about it at interview", 4 = "This was NOT shown in the mock crime, but I gave it as a lie in response to a question at interview", 5 = "This was NOT shown in the mock crime and I was NOT asked a question about it at interview." Error bars represent ±1 SE of the mean.

Table A9.1

Odds ratios (ORs) indicating the change in the count for each source option compared to the correct source option for all item types.

	Control Items	Items Lied About	Lies Reported			Items Truthfully Reported		
			LL	HLI	HLE	LL	HLI	HLE
1	0.51** [0.37, 0.69]	0.61* [0.42, 0.89]	0.73 [0.45, 1.20]	0.35** [0.21, 0.59]	0.54* [0.35, 0.83]	-	-	-
2	0.46** [0.33, 0.64]	-	1.24 [0.82, 1.89]	0.53** [0.37, 0.76]	0.73 [0.49, 1.09]	0.71* [0.51, 0.99]	0.35** [0.26, 0.46]	0.92 [0.63, 1.34]
3	-	0.65* [0.49, 0.86]	0.79 [0.49, 1.27]	0.56* [0.37, 0.85]	0.50* [0.33, 0.76]	0.75 [0.54, 1.04]	0.30** [0.21, 0.43]	0.71 [0.50, 1.01]
4	0.21** [0.13, 0.33]	0.30** [0.20, 0.45]	-	-	-	0.16** [0.10, 0.27]	0.09** [0.06, 0.14]	0.13** [0.07, 0.25]
5	0.58* [0.42, 0.81]	0.43** [0.30, 0.63]	1.19 [0.80, 1.78]	0.61* [0.39, 0.94]	1.18 [0.79, 1.74]	0.28** [0.18, 0.42]	0.07** [0.04, 0.13]	0.20** [0.12, 0.35]

Note: * indicates $p < .05$; ** indicates $p < .001$. LL = low load, HLI = high intrinsic load, HLE = high extraneous load. The source option numbers correspond to the following statements: 1 = "This was shown in the mock crime and I gave it as a truthful answer to a question at interview", 2 = "This was shown in the mock crime, but I lied when I was asked about it at interview", 3 = "This was shown in the mock crime, but I was NOT asked a question about it at interview", 4 = "This was NOT shown in the mock crime, but I gave it as a lie in response to a question at interview", 5 = "This was

NOT shown in the mock crime and I was NOT asked a question about it at interview." ORs of <1 indicate that the option was selected less often than the correct option; ORs of >1 indicate that the option was selected more often than the correct option. Values in parentheses indicate 95% confidence intervals. For control items and items lied about, the interaction between Source Option and Cognitive Load was not significant and therefore only ORs for the main effect of Source Option are reported.

Appendix 10. CIT questions asked in Session 3 of Experiment 6.

(S) indicates substitute irrelevants to replace irrelevants that participants gave as their lie in Session 2 of the experiment.

Question	Probe	Irrelevants
Was this the item used to conceal the CCTV camera?	Shaving foam	Cloth Spray paint Duct tape Blu tack Cream (S)
Was this the code for the combination lock?	8592	3546 7961 5074 5512 8651 (S)
Was this the item used to cut the padlock off the device case?	Bolt cutters	Hacksaw Wire cutters Metal snips Lock picks Axe (S)
Was this the item that the computer files were copied onto?	USB stick	Mini CD SD card Micro computer External hard drive Floppy disk (S)
Was this the phone that was used to photograph the device blueprints?	iPhone	Samsung Galaxy Sony Xperia Google Pixel Huawei Mate HTC (S)
Was this the security hologram image?	Brain	Earth Clock Jellyfish Heart Skull (S)

Was this where the device was stored?	Cupboard	Filing cabinet Shelf Locker Safe Desk (S)
Was this the object that you took the device away in?	Sports bag	Backpack Briefcase Satchel Suitcase Tote bag (S)