A Stability Bias effect amongst lie-tellers:

Testing the 'miscalibration' and 'strategic' hypotheses.

Adam Charles Harvey*,

Department of Psychology, University of Portsmouth, UK

Department of Social Sciences, University of the West of England (UWE) Bristol, UK,

Psychological Sciences Research Group (PSRG)

Aldert Vrij

Department of Psychology, University of Portsmouth, UK

Lorraine Hope

Department of Psychology, University of Portsmouth, UK

Samantha Mann

Department of Psychology, University of Portsmouth, UK

Key Words (5): Deception; stability bias effect; forgetting; delay; metacognition

Word counts: Introduction: (**1098 words**); Discussion: (**764 words**); Total excluding abstracts; tables; figures, references, and supplementary analysis: (**4312 words**)

*Correspondence concerning this article should be addressed to Adam Charles Harvey, University of the West of England (UWE) Bristol, Department of Social Sciences, Frenchay Campus, Bristol, BS16 1QY, United Kingdom (UK). Email: adam3.harvey@uwe.ac.uk

Author contribution statement

- The experiment was designed by the first (A.H.), second (A.V.) and third (L.H.) authors.
- The data was collected by the first author (A.H.).
- The manuscript was written by the first author (A.H.), with comments from the second (A.V.), third (L.H.) and fourth (S.M.) authors.
- Data analysis was conducted by the first author (A.H.), with comments from the second (A.V.) and third (L.H.) authors.

The authors declare no conflict of interest.

Abstract (150)

Unlike truth-tellers’ statements that show forgetting, lie-tellers’ statements appear less sensitive to delay. For lie-tellers this failure to correctly simulate forgetting has been referred to as a stability bias. This experiment tests two explanations for this stability bias: the ‘miscalibration’ hypothesis and the ‘strategic’ hypothesis. Using a 2 (Task Type: recall estimate vs strategic estimate) × 2 (Delay: immediate vs. three-week) design, participants (*n* = 142) either estimated how much detail a truth-teller might remember from an intelligence briefing (testing the miscalibration hypothesis), or how much detail was necessary to make a fabricated statement about the same intelligence briefing appear convincing to others (testing the strategic hypothesis). Before making these estimates, participants were informed that the briefing occurred immediately beforehand, or three-weeks beforehand. Recall estimates correctly predicted forgetting would occur after a three-week delay. Strategic estimates did not vary as a function of statement-time. No differences in subjective beliefs emerged.

General Audience Summary (141 words)

It is well-established that truth-tellers report fewer details when interviewed after a delay (compared to when interviewed immediately) due to forgetting. In contrast, lie-tellers tend to report similar amounts of detail when interviewed immediately or after a delay. Lie-tellers' failure to correctly simulate forgetting is what's known as a stability bias effect. The current experiment examines the processes underlying this stability bias effect amongst deceivers. We found that participants were sensitive to delay and predicted forgetting when estimating how much detail could be remembered. In contrast, participants estimated that similar amounts of detail would be needed to convince others that a statement was genuine. This was the case when the statement was given immediately following an event, and after a three-week delay. Results suggest that the stability bias effect amongst lie-tellers is underpinned by their motivation to appear convincing to others.

Introduction

Genuine memory performance is time sensitive (Anderson, 1983; Ayers & Reder, 1998; Wixted & Carpenter, 2007). Information recalled from memory becomes less accessible with increased delay between encoding and retrieval (Penrod, Loftus & Winkler, 1982; Wixted & Ebbesen, 1997), with declining performance occurring rapidly at first, before plateauing (a pattern known as the forgetting curve; Ebbinghaus, 1885; Murre & Dros, 2015). In contrast, when lie-tellers provide statements after delays, their reported statements tend to underestimate the effects of forgetting (Harvey et al., 2017a; 2019; 2020; Izotovas et al., 2018; Nahari, 2018). This has been referred to as a ‘stability bias’ effect (Harvey et al., 2017b). Although the mechanism responsible for this stability bias effect amongst deceivers is currently not understood (Harvey et al., 2019), two potential explanations have been proposed: (i) the ‘miscalibration’ hypothesis and (ii) the ‘strategic’ hypothesis (Harvey et al., 2017b; 2019). According to the miscalibration hypothesis, the stability bias effect is attributable to lie-tellers’ inability to accurately estimate the extent of genuine forgetting over time. According to the strategic hypothesis, the stability bias is underpinned by lie-tellers’ motivation to over-report details after delays in order to appear convincing to others. The aim of this current experiment is to test the extent to which either hypothesis can account for the stability bias effect amongst lie-tellers.

In most deception experiments, truth tellers and lie-tellers are typically interviewed immediately after experiencing some target event (Harvey et al., 2019; Nahari, 2018; Vrij, 2008). Fewer studies have explored the effects of delay upon suspects’ verbal statements (c.f. Harvey et al., 2017b; Izotovas et al., 2018; Nahari, 2018). However, from this initial research on the effects of delay, three main findings have emerged. Firstly, consistent with established memory theory predicting that forgetting occurs over time (Anderson, 1983; Ayers & Reder, 1998; Wixted & Carpenter, 2007), truth-tellers report fewer details when interviewed after delays compared to when interviewing occurs immediately (e.g. Harvey et al., 2017a; Nahari, 2018). Secondly, and the focus of the current experiment, lie-tellers fail to accurately simulate this pattern of forgetting after delays in their own statements (e.g. Izotovas et al., 2018; also see Vrij et al., 2009), thereby displaying a stability bias effect (Harvey et al., 2017b). This stability bias emerges both when lie-tellers incorporate *imagined* delays into their statements (Harvey et al., 2017b, Study 1) and when lie-tellers fabricate statements after experiencing *real* delays (Harvey et al., 2017b, Study 2; Harvey et al., 2017a; 2019, 2020). Thirdly, as truth-tellers typically report more details than lie-tellers when interviewed without delay (Amado, Arce, Fariña, & Vilarino, 2016), deceptive and genuine statements are more similar in terms of reported detail when interviewing occurs after delays, compared to when interviewing is immediate (Harvey et al., 2017a, Izotovas et al., 2018; Nahari, 2018).

Examination of the stability bias is important. For example, it may be useful to know *why* lie-tellers display a stability bias effect after delays in order to develop interviewing techniques to exploit their specific mental processes (e.g. Harvey et al., 2019). Although it is theoretically unclear why the stability bias effect amongst deceivers emerges, at least two alternative hypotheses can be posited.

The miscalibration hypothesis is theoretically based upon the dual view of metacognitive judgements derived from the Cue Utilization Approach (Koriat, 1997; Koriat et al., 2004; Kornell et al., 2009). According to this theory, metacognitive judgements are either (a) theory-based or (b) experience-based. The former requires the explicit application of theory and belief (i.e. knowledge about memory decay across time), whist the latter utilizes processing of actual items in memory (i.e. strength and precision of genuine recollection to act as a guide). Although the relative dominance of each explanation is unclear (but see Harvey et al., 2017b), it is well-established that individuals typically hold false beliefs about memory, e.g. believing that information is permanently stored in memory, in a manner similar to that of a computer or video camera (Legaut & Laurence, 2007; Loftus & Loftus, 1980; Magnussen et al., 2008; Ost, et al., 2016; Simons & Chabris, 2011). If lie-tellers hold false beliefs about general memory performance (or false beliefs regarding their own memory ability), this may lead to overestimating the appropriate level of detail to report after delays. Such an overestimation may occur when individuals make only a single estimate of recall performance (i.e. what would be recalled after a delay) in a between-subjects manner, or when individuals make several estimates of recall performance (i.e. what would be recalled after 1 week, 1 month, or 1 year) in a within-subjects manner.

In contrast, the strategic hypothesis is theoretically based upon interviewee's impression management strategies, Unlike truth tellers who tend to take their credibility as self-evident (Gilovich, Savitsky, & Medvec, 1998; Jordan & Hartwig, 2013), lie-tellers are motivated to convey an honest impression (e.g. Köhnken, 1989, 1996, 2004). Therefore, to maximise their chances of appearing convincing to others, lie-tellers may be strategically motivated to report statements rich in detail (Hartwig et al., 2007; Masip & Herrero, 2013; Strömwall et al., 2006). This strategy makes good practical sense as detailed statements are more likely to be judged as credible compared to sparsely detailed statements (Bell & Loftus, 1989; Johnson, 2006; Johnson, Foley, Suengas, & Raye, 1988). This hypothesis also accounts for why the stability bias effect generalises across settings whereby lie-tellers (i) incorporate an imagined delay into their statement (Harvey et al., 2017b, Study 1), and (ii) lie after experiencing a real delay (Harvey et al., 2017b, Study 2; Harvey et al., 2017b; 2019, 2020). In both settings, lie-tellers are concerned with appearing as credible, not estimating the extent of genuine recall, after delays.

The strategic explanation is further supported by a close examination of the methodology (specifically, the pre-interview instructions) used in previous deception research. For example, Harvey et al. (2017b, Study 1) informed lie-tellers that they were free to report any information necessary to 'convince the interviewer' that they were actually genuine (also see Harvey et al., 2017a; 2019; 2020). Analogous instructions have been used by other researchers (Izotovas et al., 2018; Nahari, 2018). Thus, the strategic explanation for the stability bias effect amongst deceivers is also consistent with lie-tellers closely following experimental instructions.

Based upon the previous theoretical considerations, we expected that when individuals made recall estimates (estimating how much detail could actually be recalled as a function of delay), declining performance over time would be predicted (Hypothesis 1). In contrast, we expected that when individuals made strategic estimates (estimating how much detail is required for a statement to appear convincing as a function of delay), a 'stability bias' pattern would be predicted (Hypothesis 2).

Method

Design

A 2 (Task Type: Recall Estimate vs Strategic Estimate) × 2 (Delay: Immediate vs. Three-week delay) between-subject experimental design was used. The main dependent measures were the estimated percentage of details (e.g. 50%) required for the recall or strategic task, and responses on the post-experiment questionnaire responses. Our post-experiment questionnaire included four questions assessing participants' beliefs about memory from the Memory Assessment Questionnaire (Ost et al., 2013; also see Ost et al., 2017). These four items constitute the entire 'mailability of memory' component of the Memory Assessment Questionnaire (Ost et al., 2017), with low scores indicating a belief that memories are accurately stored and retrievable; that "true" and "false" memories can be reliably distinguished. This included questions such as, '*…on a scale of 1 (strongly agree) to 4 (strongly disagree), to what extent do you agree with the following statement: Memory is* *like a computer, accurately recording events as they actually occurred…*’). Our questionnaire also included five questions assessing participants’ beliefs about their own metacognitive ability from the Squire Subjective Memory Questionnaire (SSMQ; see van Bergson et al., 2010). These five items were selected from the eighteen SSMQ items (van Bergson et al., 2010), with low scores indicating a belief that memory performance is poor. This included questions such as, ‘*…on a scale of -4 (disastrous) to 4 (perfect), how would you rate the following statement: My ability to recall things when I really try is…*’).

Ethics

A favorable ethical review decision was given, prior to the research, by the Science Faculty Ethics Committee (SFEC 2018-018), University of Portsmouth (UK). Our Institution’s SFEC conforms to the British equivalent of APA ethical standards (see British Psychological Society, 2009; 2014; 2018). All participants’ rights were upheld during the research.

Participants

An a priori power analysis using G*Power (Faul, Erdfelder, Lang & Buchner, 2007; Faul, Erdfelder, Buchner & Lang, 2009), assuming a medium effect size of $f = 0.25$ ($\alpha = 0.05$) for four groups, indicated a sample size of 128 would be sufficient for power of 0.80 (Cohen, 1988, 1992). However, for tests that examine interaction effects (e.g. the Task Type × Delay interaction effect explored in the current experiment), G*Power tends to underestimate the number of required participants to achieve 80% power (for more information, see https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/). To account for this, and compensate for any potential participant attrition (i.e. participants not following experimental instructions and requiring exclusion), an additional 14 individuals were recruited (approximately 10% of the original sample size estimate). A total of 142 participants were recruited on a voluntarily basis, comprising 111 females and 31 males, aged between 18 and 68 years (*M* = 24.87 years, *SD* = 10.64, 95% CI [23.27, 26.91]), from the undergraduate (*n* = 106), postgraduate (*n* = 24) and staff (*n* =12) communities.

Procedure

Participants were recruited via adverts on the university's online participant recruitment platform and posters placed around the campus. The adverts solicited individuals to participate in a experiment on 'memory performance' in intelligence gathering settings. Individuals who had previously taken part in similar previous research were not eligible to participate.

All participants arrived individually at the laboratory at pre-arranged times. Following consent procedures, participants were then randomly allocated to one of four experimental conditions: the immediate (*n* = 35) or delayed (*n* = 35) Recall Estimate task condition, or the immediate (*n* = 36) or delayed (*n* = 36) Strategic Estimate task condition. All participants were provided with experimental instructions (specific to their condition) and a detailed statement (identical for all participants).

Experimental instructions

Participants in the Recall Estimate task conditions (*n* = 71) were instructed to imagine themselves as an intelligence agent who attended a classified briefing. Participants were informed that they would be provided with a written statement accurately describing all the information disclosed during the classified briefing. Participants were further informed that their task was to estimate how many of these details they, as an agent present at the briefing, could correctly recall.

Participants in the Strategic Estimate task condition (*n* = 71) were instructed to imagine themselves as an intelligence agent who attended a classified briefing. However, they were then instructed that their mission was to mislead a hostile intelligence agency by passing on false information about this classified briefing. Participants were informed that they would be provided with a written statement outlining all the false information that had been prepared to mislead the hostile agency.

Briefing Statement

All participants were then provided with the same detailed briefing statement that was 218 words long and adapted from a video stimulus used in previous research (Harvey et al., 2017a; 2017b, 2019, 2020). Using Reality Monitoring criteria (see Vrij, 2008), each spatial detail (information about locations and spatial arrangements of people and/ or objects; n = 19); temporal detail (information about when an event happened or explicit descriptions of sequences of events; n = 6); and perceptual detail (information regarding sounds, smells, tastes, physical sensations, and visual details; n = 55) within the briefing statement was underlined for the participants (total details: n = 80). RM criteria are used extensively in research due to (a) their strong theoretical foundation; (b) their ability to accurately index the level of detail contained within a statement; and (c) their detailed and clearly operationalised criteria that facilitates good inter-rater agreement in scoring (Masip et al., 2005; Vrij, 2008). Note. The RM coding for this briefing statement can be found in the supplementary materials on the first author's OSF account (https://osf.io/yd8hf/).

In the Recall Estimate condition, participants were informed that their task was to estimate how many of these details (from the detailed statement) they, as the agent, could correctly recall. Participants in the immediate condition (n = 35) were instructed to estimate how many details they could correctly recall *immediately* after the briefing, whereas participants in the delayed condition (n = 36) estimated how many details they could correctly recall *three-weeks after* the briefing.

In the Strategic Estimate condition, participants were informed that their task was to estimate how many of these false details (from the detailed statement) they, as the agent, should report to convince others in the hostile intelligence agency that they are being honest. Participants in the immediate condition (*n* = 35) were instructed to estimate how many details they should report to appear convincing immediately after the briefing, while participants in the delayed condition (*n* = 36) estimated how many details they should report to appear convincing three weeks after the briefing.

All participants reported their estimation in the same way. Participants provided a percentage estimate (e.g. 50%) of the amount of detail they believed *would* be recalled (Recall Estimate condition) or *should be* reported (Strategic Estimate condition). After providing their estimate, all participants were provided with an identical post-experiment questionnaire.

Post-experiment questionnaire

The post-experiment questionnaire comprised three sections: The first section assessed participant's level of motivation during the experiment, their engagement with the experimental task, and four manipulation checks. The second section comprised nine questions assessing participant beliefs about memory in general (metamemory), or beliefs regarding their own memory performance (metacognition). See Supplementary analysis for descriptive statistics. In the final section, all participants read a statement similar in content to the first (i.e. also about a 'classified intelligence briefing'). To check whether participants in both Task conditions could predict forgetting occurs over time, all participants (regardless of the experimental conditions) were instructed to estimate what percentage of the (second) statement's details they would recall correctly after (i) one day, (ii) one month, and (iii) one year. In this manner, participants estimated their recall in a within-subjects manner. Note. The post-experiment questionnaire can be found in the supplementary materials on the first author's OSF account (https://osf.io/yd8hf/).

This second statement was 154 words long, composed of 40 RM details (n = 23 perceptual details; n = 11 spatial details; n = 6 temporal details). The RM coding for this second statement can be found in the supplementary materials on the first author's OSF account (https://osf.io/yd8hf/).

Once participants finished the post-experiment questionnaire, the experiment finished. Participants were thanked, debriefed and left the laboratory.

Results

Analysis plan

For parsimony we report Cohen's f for all ANOVA effect sizes (whereby f = 0.1, 0.25, and 0.4 correspond to small, medium and large effects respectively; Cohen, 1988). Following recommendations (Cohen, 1988; Lakens, 2013), for all other contrasts we report Cohen's d (with 95% confidence intervals).

To assess the strength of evidence, and in addition to null hypothesis significance testing, we also calculated a Bayes Factor (BF) score (e.g. Wagenmakers et al., 2016) using a default Bayesian t test (with the default Cauchy's prior of 0.707; see Lakens, 2016) and open-source JASP software (https://jasp-stats.org see Wagenmakers et al., 2017b). BF_{10} is the Bayes factor giving the evidence for an alternative hypothesis over the null (and increases when evidence more strongly supports the alternative hypothesis). BF_{01} is the Bayes factor giving the evidence for the null hypothesis over the alternative (and increases when evidence more strongly supports the null hypothesis). *Note*: $BF_{10} = 1/ BF_{01}$.

Motivation

Overall, interviewees' reported motivation was high (overall M = 6.41, SD = 1.03, 95% CI [6.23, 6.57]). No significant main or interaction effects emerged, all f < 0.09, all p's

> .260, BF_{01} 4.35 – 20.42. This suggests that participants in all conditions were equally motivated to perform well during the experiment.

Engagement

Overall, interviewees' reported engagement with the instructions was high (overall M = 6.24, SD = 0.95, 95% CI [6.08, 6.39]). No significant main or interaction effects emerged, all $f < 0.07$, all p's > .368, BF_{01} 4.68 – 33.33. This suggests participants in all conditions were equally engaged with reading and understanding the experimental instructions during the experiment.

Manipulation checks

We conducted four 2 (Task Type: Recall Estimate vs. Strategic Estimate) × 2 (Statement Time: Immediate vs Three-week delay) between-subjects analyses of variance (ANOVAs) with four manipulation check items as dependent variables (all checking if participants followed condition-specific instructions). For these analyses, see Supplementary analysis. Collectively, these analyses support the validity of our Task Type and Statement Time manipulations.

Hypothesis Testing

We conducted a 2 (Task Type: Recall Estimate vs. Strategic Estimate) × 2 (Delay: Immediate vs Three-week delay) between-subjects ANOVAs to examine the estimated percentage of details to include to appear convincing.

A significant main effect for Task Type was found for the estimated percentage of details to include, $F(1, 138) = 47.78$, $p < .001$, $f = 0.59$, $BF_{10} = 1.17$^9, such that participants in the Strategic Estimate condition estimated including a higher percentage of details than participants in the Recall Estimate condition. A significant main effect was found for Statement Time, $F(1, 138) = 9.86$, $p = .002$, $f = 0.27$, $BF_{10} = 1289.34$, such that participants in the Immediate condition estimated including a higher percentage of details than participants in the Three-week delay condition. Furthermore, a significant Task Type × Statement Time interaction effect was also found, $F(1, 138) = 16.19$, $p < .001$, $f = 0.34$, $BF_{10} = 815.63$.

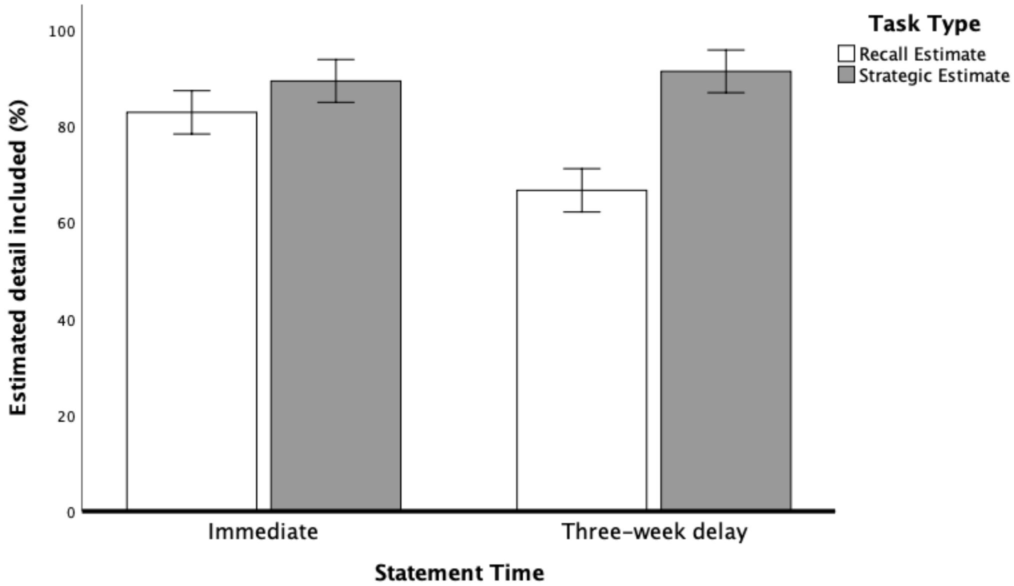

Figure 1: Participants' estimates of the percentage of details to be included, as a function of Task Type and Statement Time (with 95% error bars).

Participants in the Recall Estimate condition included a greater percentage of details in the Immediate condition ($M = 82.86$, $SD = 11.33$, 95% CI [78.95, 86.46]) than those in the Three-week delayed condition ($M = 66.66$, $SD = 15.46$, 95% CI [61.70, 71.56]), $t(62.34) = 5.00$, $p < .001$ (one-tailed), $d= 1.20$, 95% CI [0.68, 1.70]. Bayesian analysis showed our data supported the alternative hypothesis, $BF_{10} = 3656.97$. These findings support Hypothesis 1. In contrast, in the Strategic condition, no significance difference for the percentage of details included emerged between participants in the Immediate condition ($M = 89.39$, $SD = 15.31$, 95% CI [83.83, 94.00]) compared to those in the Three-weeks delayed condition ($M = 91.39$, $SD = 11.16$, 95% CI [86.32, 94.66]), $t(70) = 0.63$, $p = .528$ (two-tailed), $d= 0.15$, 95% CI [-

0.31, 0.61]. Bayesian analysis showed our data supported the null hypothesis, $BF_{01} = 3.46$. These findings support Hypothesis 2.

Individual's beliefs about memory

To examine participants' beliefs about memory as a function of our experimental conditions, we conducted a 2 (Task Type: Recall Estimate vs. Strategic Estimate) × 2 (Statement Time: Immediate vs. Three-week delay) between-subjects ANOVA using the combined mean score of the four question items (from the Memory Assessment Questionnaire; Ost et al., 2013; 2017) as the dependent measure. No significant main or interaction effects emerged, all $f < 0.04$, all p's > .658, BF_{01} 7.20 – 40.27. Thus, our participants did not differ in beliefs about memory performance (i.e. metamemory). For descriptive statistics of the individual question items, see Supplementary analysis.

Individual's beliefs about their own memory ability

To examine participants' ratings of their own memory ability as a function of our experimental conditions, we conducted a 2 (Task Type: Recall Estimate vs. Strategic Estimate) × 2 (Statement Time: Immediate vs. Three-week delay) between-subjects ANOVA using the combined sum total score of the five question items (from the SSMQ; see van Bergson et al., 2010) as dependent measure. No significant main or interaction effects emerged, all $f < 0.02$, all p's > .842, BF_{01} 7.91 – 42.26. Thus, our participants did not differ in beliefs regarding their own memory ability (i.e. metacognition). For descriptive statistics, see Supplementary analysis.

Within-subject estimates of recall

To check whether participants in both Task Type conditions could predict that forgetting occurs over time, we conducted a 2 (Task Type: Recall Estimate vs. Strategic Estimate) × 2 (Statement Time: Immediate vs. Three-week delay) × 3 (Retention Interval: 1 day vs. 1 month vs. 1 year) mixed ANOVA). Task Type and Statement Time were the between-subject factors and Retention Interval was the within-subjects factor. Mauchly's Test indicated that the assumption of sphericity was violated, $\chi^2(2) = 0.67$, $p < .001$. As a result, Huynd-Feldt corrections were applied.

A significant main effect of Retention Interval emerged, $F(1.55, 213.41) = 335.83$, $p < .001$, $f = 1.56$, $BF_{10} = 2.55$^13. Post-hoc LSD tests revealed that participants estimated a higher percentage of recall after one day ($M = 88.27$, $SD = 17.40$, 95% CI [85.36, 91.19]) compared to after one month ($M = 55.28$, $SD = 21.02$, 95% CI [51.78, 58.78]) or after one year ($M = 37.98$, $SD = 24.96$, 95% CI [33.81, 42.15]) (all p's < .001). No other significant main or interaction effects emerged, all $f < 0.09$, all p's > .283, BF_{01} 8.14– 11641.22. 40.27. Thus, our participants did not differ in their ability to predict (in a within-subjects manner) that forgetting occurs over time.

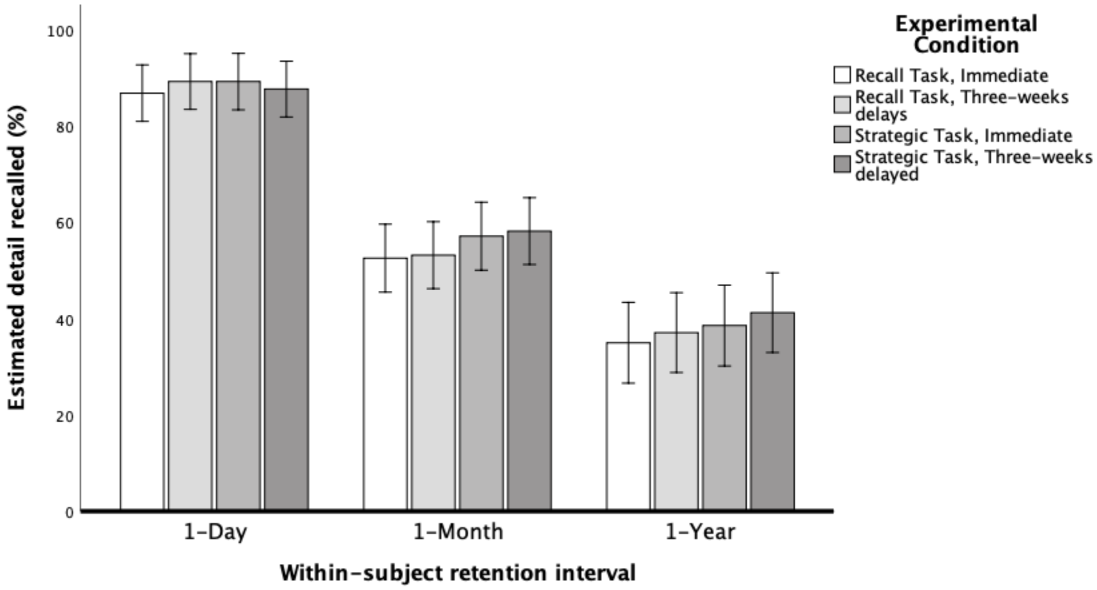

Figure 2: Participants’ estimates of the percentage of details recalled, as a function of Task Type, Statement Time, and Retention Interval (with 95% error bars).

Discussion

The aim of this current experiment is to test the extent to which either the miscalibration hypothesis or strategic hypothesis can account for the stability bias effect amongst lie-tellers. Participants in the Recall Estimate condition (estimating how much detail the agent *could* recall) correctly predicted the likelihood of declining memory performance over time (i.e. forgetting) when estimating how much detail to include. These results support Hypothesis 1. In contrast, participants in the Strategic Estimate condition (estimating how much detail the agent *should* report to appear convincing) reported similar estimates of detail to include to appear convincing irrespective of the statement being provided immediately or after a three-week delay. These results support Hypothesis 2. Accordingly, the strategic hypothesis (c.f. miscalibration hypothesis) appears to better account for the stability bias effect amongst lie-tellers (e.g. Harvey et al., 2017a; 2017b; 2019).

No differences emerged for participants' beliefs about general memory performance (metamemory), or beliefs about their own memory ability (metacognition). Further no differences emerged for participants' objective ability to predict their own (declining) recall performance over time. Therefore, our participants' apparent insensitivity to the effects of delay in the Strategic (but not Recall) condition cannot be attributed to differences in beliefs about their own or others' memories.

Our results suggest that, despite an ability to correctly predict that forgetting occurs, individuals discount such information when judging how much detail should be reported after delays to appear convincing to others. These findings make theoretical sense. According to the strategic hypothesis, unlike truth-tellers who tend to take their credibility as self-evident (Gilovich, Savitsky, & Medvec, 1998; Jordan & Hartwig, 2013), deceptive individuals must project a credible impression to others to avoid detection (e.g. Köhnken, 1989, 1996, 2004). Therefore, to maximise their chances of appearing credible to others (Hartwig et al., 2007; Masip & Herrero, 2013; Strömwall et al., 2006), lie-tellers may over-report details after delays. This strategy may be effective as detailed statements are more likely to be judged as credible compared to sparsely detailed statements (Bell & Loftus, 1989; Johnson, 2006; Johnson, Foley, Suengas, & Raye, 1988). Thus, the stability bias effect observed amongst deceivers (e.g. Harvey et al., 2017a; 2017b; 2019; 2020) appears to be a by-product of lie-tellers' (i) motivation to appear credible to others, and (ii) seemingly static beliefs about what constitutes a credible impression. However, creating a credible impression is typically believed by both lie-tellers and truth-tellers to involve more than just providing richly detailed statements (Hines et al., 2010). Hence, other dimensions of verbal content (e.g. providing clear, careful stories lacking contradictions) may be manipulated by lie-tellers after delays in a similar manner to overall reported detail (also see DePaulo et al., 2003). Future research should explore this possibility.

It might be argued that our experiment activated participants’ metacognitive knowledge (that forgetting occurs over time) in a manner that is unlikely to translate to laboratory-based lie-detection research or the real world. That is, we elicited an explicit estimate of recall after delays. This may have activated our participants’ knowledge of forgetting, allowing them to predict a decline in performance over delays (for a similar argument, see Koriat et al., 2004). However, when statements (rather than explicit estimates of recall performance) are elicited in more typical laboratory research (e.g. Harvey et al., 2017b; Izotovas et al., 2018; Nahari, 2018), the same knowledge of forgetting may not be activated. Participants in such research may be unable to predict forgetting. We believe this is unlikely as previous research has shown that, under certain conditions, individuals can quite accurately predict forgetting (Koriat et al., 2004). Furthermore, this argument does not explain our experiment’s core finding that estimates of how much detail is needed to appear convincing did not vary as a function of delay, whereas estimates concerning recall did. Rather, we believe it likely that lie-tellers can (even in laboratory research) predict that truth-tellers forget information. However, this is unlikely to be their primary concern when making a fabricated statement (or preparing a lie-script). These topics warrant further investigation.

Individuals in our experiment did not freely generate their own statements, but were given information and asked to make estimates regarding that information. Individuals were thus provided with a ‘lie script’ (Colwell et al., 2007) or ‘self-manipulated memory’ (Nahari, 2018) that served as the basis for their estimates. It is possible that different results may have been obtained if individuals were free to generate their own lie-script (or self-manipulated memory) to appear convincing. Future research could explore this possibility.

Conclusion

In sum, our results provide compelling support for the strategic explanation of the stability bias effect amongst deceivers.

References

Amado, B. G., Arce, R., Farina, F., & Vilariño, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, *16*(2), 201-210. https://doi.org/10.1016/j.ijchp.2016.01.002

Anderson, J. R. (1983a). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 261-295. https://doi.org/10.1016/S0022-5371(83)90201-3

Anderson, J. R. (1983b). *The architecture of cognition.* Hillsdale, NJ: Erlbaum.

Ayers, M. S. & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*(1), 1-21. https://doi.org/10.3758/BF03209454

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology*, *56*(5), 669-679. https://doi.org/10.1037/0022-3514.56.5.669

British Psychological Society (2009). *Code of Ethics and Conduct. Guidance by the Ethics Committee of the British Psychological Society*. Leicester, UK: British Psychological Society.

British Psychological Society (2014). *Code of Ethics and Conduct. Guidance by the Ethics Committee of the British Psychological Society*. Leicester, UK: British Psychological Society.

British Psychological Society (2018). *Code of Ethics and Conduct. Guidance by the Ethics Committee of the British Psychological Society*. Leicester, UK: British Psychological Society.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment criteria indicative of deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling*, *4*(3), 167-180. https://doi.org/10.1002/jip.73

Cumming, G. (2014). The new statistics: Why and how. *Psychological science*, *25*(1), 7-29, https://doi.org/10.1177/0956797613504966

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. https://doi.org/10.1037/0033-2909.129.1.74

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. Leipzig: Duncker & Humblot.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

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. https://doi.org/10.3758/BF03193146 .

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. Journal of Personality and Social Psychology, 75(2), 332–346. https://doi.org/10.1037/0022-3514.75.2.332

Harvey, A. C., Vrij, A., Hope, L., Leal, S., & Mann, S. (2017a). A stability bias effect among deceivers. *Law and human behavior, 41*(6), 519-529, http://dx.doi.org/10.1037/lhb0000258

Harvey, A. C., Vrij, A., Leal, S., Hope, L., & Mann, S. (2017b). Deception and decay: Verbal lie detection as a function of delay and encoding quality. *Journal of Applied Research in Memory and Cognition, 6*(3), 306-318, https://doi.org/10.1016/j.jarmac.2017.04.002

Harvey, A. C., Vrij, A., Leal, S., Hope, L., & Mann, S. (2019). Amplifying deceivers’ flawed metacognition: Encouraging disclosures after delays with a model statement. *Acta Psychologica, 200*, 102935. https://doi.org/10.1016/j.actpsy.2019.102935

Harvey, A. C., Vrij, A., Leal, S., Deeb, H., Hope, L., & Mann, S. (2020). Amplifying recall after delays via initial interviewing: Inoculating truth-tellers' memory as a function of encoding quality. *Acta Psychologica, 209*, 103130. https://doi.org/10.1016/j.actpsy.2020.103130

Hines, A., Colwell, K., Hiscock-Anisman, C., Garrett, E., Ansarra, R , & Montalvo, L. (2010). Impression management strategies of deceivers and honest reporters in an investigative interview. *European Journal of Psychology Applied to the Legal Context 2*(1). 73-90.

Izotovas, A., Vrij, A., Hope, L., Mann, S., Granhag, P. A., & Strömwall, L. A. (2018). Facilitating memory-based lie detection in immediate and delayed interviewing: The role of mnemonics. *Applied Cognitive Psychology*, *32*(5), 561-574. https://doi.org/10.1002/acp.3435

Johnson, M. K. (2006). Memory and reality. *American Psychologist*, *61*(8), 760-771. https://doi.org/10.1037/0003-066X.61.8.760

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. https://doi.org/10.1037/0096-3445.117.4.371

Jordan, S., & Hartwig, M. (2013). On the phenomenology of innocence: The role of belief in a just world. *Psychiatry, Psychology and Law*, *20*(5), 749-760. https://doi.org/10.1080/13218719.2012.730903

Köhnken G. (1989) Behavioral Correlates of Statement Credibility: Theories, Paradigms, and Results. In: Wegener, H., Lösel, F., Haisch, J. (Eds.), Criminal Behavior and the Justice System. *Research in Criminology* (pp. 271-289). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-86017-1_18

Köhnken, G. (1996). Social psychology and the law. In: Semin, G.R., & Fiedler, K. (Eds.), *Applied social psychology* (pp. 257–281). Sage Publications, Inc. https://doi.org/10.4135/9781446250556.n10

Köhnken, G. (2004). Statement Validity Analysis and the 'detection of the truth' In: Granhag, P.-A., & Strömwall, L. (Eds.), *The detection of deception in forensic contexts* (pp. 41–63). Cambridge University Press. https://doi.org/10.1017/CBO9780511490071.003

Koriat, A. (1997). Monitoring one's own knowledge during Experiment: A cue-utilization approach to judgments of learning. *Journal of experimental psychology: General*, *126*(4), 349-370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643-656. https://doi.org/10.1037/0096-3445.133.4.643

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, *138*(4), 449-468. https://doi.org/10.1037/a0017350

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D. (2016). Power analysis for default Bayesian t-tests [Blog post]. *The 20% statistician.* Retrieved from https://daniellakens.blogspot.com/2016/01/power-analysis-for-default-bayesian-t.html

Legault, E., & Laurence, J. R. (2007). Recovered memories of childhood sexual abuse: Social worker, psychologist, and psychiatrist reports of beliefs, practices, and cases. *Australian Journal of Clinical and Experimental Hypnosis*, *35*(2), 111-133.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35*(5), 409–420. https://doi.org/10.1037/0003-066X.35.5.409

Magnussen, S., & Melinder, A. (2012). What psychologists know and believe about memory: A survey of practitioners. *Applied Cognitive Psychology, 26*(1), 54-60. https://doi.org/10.1002/acp.1795

Masip, J., & Herrero, C. (2013). ‘What Would You Say if You Were Guilty?’Suspects' Strategies During a Hypothetical Behavior Analysis Interview Concerning a Serious Crime. *Applied cognitive psychology*, *27*(1), 60-70. https://doi.org/10.1002/acp.2872

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law, 11*(1), 99-122. https://doi.org/10.1080/10683160410001726356

Murre, J. M., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PloS one*, *10*(7), e0120644. http://dx.doi.org/10.1371/journal.pone.0120644

Nahari G. (2018). Reality monitoring in the forensic context: Digging deeper into the speech of liars. *Journal of Applied Research in Memory and Cognition*, *7*(3), 432-440. https://doi.org/10.1016/j.jarmac.2018.04.003

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*(2), 241–301. https://doi.org/10.1037/1082-989X.5.2.241

Ost, J., Easton, S., Hope, L., French, C. C., & Wright, D. B. (2017). Latent variables underlying the memory beliefs of Chartered Clinical Psychologists, Hypnotherapists and undergraduate students. *Memory, 25*(1), 57-68. https://doi.org/10.1080/09658211.2015.1125927

Ost, J., Wright, D. B., Easton, S., Hope, L., & French, C. C. (2013). Recovered memories, satanic abuse, dissociative identity disorder and false memories in the UK: A survey of clinical psychologists and hypnotherapists. *Psychology, Crime & Law, 19*(1), 1-19. https://doi.org/10.1080/1068316X.2011.598157

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the US population. *PloS one*, *6*(8).

Strömwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law*, *12*(2), 207-219. https://doi.org/10.1080/10683160512331331328

van Bergen, S., Brands, I., Jelicic, M., & Merckelbach, H. (2010). Assessing trait memory distrust: Psychometric properties of the Squire Subjective Memory Questionnaire. *Legal and Criminological Psychology, 15*(2), 373-384. https://doi.org/10.1348/135532509X471960

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

Vrij, A., Leal, S., Granhag, P.A., Mann, S., Fisher, R.P., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior, 33*(2), 159-166. https://doi.org/10.1007/s10979-008-9143-y

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Meerhoff, F. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*, 58-76, https://doi.org/10.3758/s13423-017-1323-7

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Matzke, D. (2017). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review, 25*(1), 35-57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E. J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169-176, https://doi.org/10.1177/0963721416643289

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432. https://doi.org/10.1037/a0022790

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*, *18*(2), 133-134. https://doi.org/10.1111/j.1467-9280.2007.01862

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: Quantitative analysis of individual subject forgetting functions. *Memory & Cognition, 25*(5), 731-739. https://doi.org/10.3758/BF03211316

Supplementary analysis

Manipulation checks

We conducted four 2 (Task Type: Recall Estimate vs. Strategic Estimate) × 2 (Statement Time: Immediate vs Three-week delay) between-subjects analyses of variance (ANOVAs) with four manipulation check items. There was a significant main effect for Task Type for our first manipulation check ('…*When making your estimates, to what extent did you consider how much detail you thought the agent could remember...*'). Participants in the Recall Estimate condition provided higher ratings (indicating they considered, to a greater extent, how much detail could be recalled) than participants in the Strategic condition (Table 1). No other significant main or interaction effects emerged, all $f < 0.13$, all p's > .143, BF_{01} 2.52 – 2.80.

As Table 1 also shows, a significant main effect for Task Type was found for our second manipulation check ('…*When making your estimates, to what extent did you consider how much detail you thought the agent should report to appear convincing*…'). Participants in the Strategic Estimate condition provided higher ratings (indicating they considered, to a greater extent, how much detail would be needed to appear convincing) than participants in the Recall Estimate condition (Table 1). No other significant main or interaction effects emerged, all $f < 0.09$, all p's > .306, BF_{01} 3.39 – 4.12. Collectively, both findings support the validity of our Task Type manipulation by showing participants in their respective conditions understood their core task differently (i.e. to estimate detail that could be recalled or would be needed to appear convincing) but consistently with instructions.

A significant main effect for Statement Time emerged for our third manipulation check ('…*When making your estimates, to what extent did you imagine that the intelligence briefing occurred immediately beforehand*…'), $F(1, 138) = 473.50$, $p < 0.001$, $f = 1.85$, BF_{10}

= 1.94^14. Participants in the Immediate condition reported imagining that the intelligence briefing occurred immediately beforehand to a greater extent (M = 6.01, SD = 0.93, 95% CI [5.80, 6.22]) than participants in the Three-weeks delayed condition (M = 2.37, SD = 1.06, 95% CI [2.11, 2.63]). As Table 1 shows, no main effect for Task Type emerged. Furthermore, the Task Type × Statement Time interaction effect was not significant, $F(1, 138) = 1.51$, $p = .221$, $f = 0.10$, $BF_{01} = 3.30$

A significant main effect for Statement Time emerged for our fourth manipulation check (‘…*When making your estimates, to what extent did you imagine that the intelligence briefing occurred three-weeks beforehand*…’), $F(1, 138) = 358.21$, $p < 0.001$, $f = 1.61$, BF_{10} = 1.20^14. Participants in the Three-weeks delay condition reported imagining that the intelligence briefing occurred three-weeks beforehand to a greater extent (M = 6.01, SD = 1.25, 95% CI [5.74, 6.27]) than participants in the Immediate condition (M = 2.13, SD = 1.18, 95% CI [1.85, 2.40]). As Table 1 also shows, no main effect for Task Type emerged. Furthermore, the Task Type × Statement Time interaction effect was not significant, $F(1, 138) = 0.07$, $p = .790$, $f = 0.02$, $BF_{01} = 6.49$. Collectively, both findings support the validity of our Statement Time manipulation by showing that participants understood their core task differently (making estimates based upon no delay or a three-week delay) but consistently with instructions.

Table 1

ANOVA Main Effects of Task Type for Manipulation Check Scores.

	Task Type							
	Recall Estimate		Strategic Estimate					
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*F-statistic*	*p-value*	*Cohen's f*	BF
Manipulation check 1: When making your estimates, to what extent did you consider how much detail you thought the agent could remember	5.99 (0.99)	5.75, 6.22	1.81, (0.94)	1.58, 2.02	$F(1, 138) = 669.14$	< .001	2.20	$BF_{10} = 2.22 \times 10^{14}$.
Manipulation check 2: When making your estimates, to what extent did you consider how much detail you thought the agent should report to appear convincing	1.87 (0.76)	1.69, 2.04	6.15, (0.88)	5.93, 6.37	$F(1, 138) = 955.20$	< .001	2.63	$BF_{10} = 1.33 \times 10^{14}$
Manipulation check 3: When making your estimates, to what extent did you imagine that the intelligence briefing occurred immediately beforehand	4.21 (2.16)	3.67, 4.70	4.17 (2.02)	3.71, 4.64	$F(1, 138) = 0.08$	.777	0.02	$BF_{01} = 5.49$
Manipulation check 4: When making your estimates, to what extent did you imagine that the intelligence briefing occurred three-weeks beforehand	4.09 (2.33)	3.51, 4.62	4.01 (2.28)	3.48, 5.59	$F(1, 138) = 0.02$	.883	0.01	$BF_{01} = 6.76$

Table 2

Participants' Beliefs about Memory from the Memory Assessment Questionnaire (Ost et al., 2013; 2017) and the SSMQ (van Bergen et al., 2010, as a Function of Experimental Conditions (Task Type and Statement Time)

	Recall Estimate task				**Strategic Estimate task**			
	Immediate		Three-week delay		Immediate		Three-week delay	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Memory Assessment Questionnaire items (Ost et al., 2013; 2017)								
Memory is like a computer, accurately recording events as they actually occurred.	1.14[a] (0.43)	1.03, 1.31	1.17[a] (0.71)	1.00, 1.44	1.11[a] (0.40)	1.00, 1.26	1.17[a] (0.65)	1.00, 1.39
Early memories, from the first year of life, are accurately stored and retrievable.	1.83[a] (0.66)	1.61, 2.06	1.77[a] (0.65)	1.57, 1.98	1.83[a] (0.77)	1.61, 2.11	1.80[a] (0.47)	1.65, 1.95

Memory is not inferenced by suggestion.	1.11[a] (0.53)	1.00, 1.31	1.11[a] (0.53)	1.00, 1.31	1.14[a] (0.54)	1.00, 1.33	1.11[a] (0.52)	1.00, 1.32
It is possible for an individual to distinguish between ‘true’ and ‘false’ memories.	1.49[a] (0.74)	1.27, 1.75	1.43[a] (0.78)	1.16, 1.69	1.50[a] (0.88)	1.23, 1.81	1.39[a] (0.73)	1.18, 1.62
SSMQ items (van Bergen et al., 2010)								
My ability to recall things when I really try is…	1.00[a] (1.95)	0.29, 1.67	0.89[a] (1.78)	0.26, 1.47	1.11[a] (1.62)	0.58, 1.59	1.22[a] (1.24)	0.56, 1.60
My ability to reach back into memory and recall what happened a few minutes ago is…	1.40[a] (1.87)	0.77, 2.03	1.23[a] (1.55)	0.70, 1.76	1.31[a] (1.45)	0.81, 1.81	1.22[a] (1.99)	0.59, 1.82
If I was asked a month from now, my ability to recall facts about this questionnaire would be…	0.20[a] (1.95)	-0.51, 0.83	0.34[a] (1.86)	-0.33, 0.89	0.17[a] (1.95)	-0.47, 2.36	0.31[a] (1.95)	-0.39, 0.94

My ability to remember things that happened more than a year ago is…	0.97[a] (2.13)	0.21, 1.68	0.74[a] (1.74)	0.14, 1.33	0.92[a] (1.81)	0.29, 1.52	0.78[a] (2.23)	0.03, 1.44
My ability to recall things that happened a long time ago is…	0.83[a] (1.95)	0.19, 1.46	0.97[a] (1.98)	0.26, 1.59	1.06[a] (1.85)	0.45, 1.67	0.97[a] (2.08)	0.29, 1.62

Note. superscripts denote significant differences ($p < .05$) between experimental conditions (only cells with different superscripts differ significantly).