APPLIED COGNITIVE PSYCHOLOGY
*Appl. Cognit. Psychol.* **19**: 985–1001 (2005)
Published online 16 June 2005 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1139

# Discriminating Adults' Genuine, Imagined, and Deceptive Accounts of Positive and Negative Childhood Events

AMANDA J. BARNIER[1]*, STEFANIE J. SHARMAN[1], LISA MCKAY[1] and SIEGFRIED L. SPORER[2]

[1]*University of New South Wales, Australia*
[2]*University of Giessen, Germany*

SUMMARY

We examined the qualitative characteristics of genuine, imagined, and deceptive accounts of positive and negative childhood events. We investigated whether trained raters could discriminate between these accounts using the Memory Characteristics Questionnaire (MCQ; Johnson, Foley, Suengas, & Raye, 1988) and the Aberdeen Report Judgment Scales (ARJS; S. L. Sporer, paper presented at the biennial meeting of the American Psychology-Law Society in Redondo Beach, California, March 1998). Participants generated three accounts. The first account was of an event that participants genuinely experienced in childhood. The second account was of an event that participants did not experience, but merely imagined happened in childhood. The third account was of an event that participants did not experience, but wrote a deceptive account to convince someone else that the event really happened in childhood. Half our participants wrote about positive events and half wrote about negative events. Ratings made by two trained judges indicated that genuine, imagined, and deceptive accounts were qualitatively different on both the MCQ and ARJS. Moreover, based on the MCQ and ARJS scores raters could discriminate whether the events had been genuinely experienced. We discuss the theoretical and practical implications of these findings. Copyright © 2005 John Wiley & Sons, Ltd.

Historically, psychological research on the detection of deception has focused on the non-verbal and physiological cues associated with deception, such as avoiding eye contact and measuring galvanic skin response (Kapardis, 1997). However, lie detection using non-verbal cues is usually little better than chance. For example, Vrij's (2000) review of approximately 40 studies found that the accuracy rate for detecting lies was only 44%. More recent research has focused on verbal cues; that is, the cues found in the content of people's statements. In this experiment, we compared the ability of two content analysis tools to discriminate between accounts of genuinely experienced events, accounts of imagined events, and deceptive accounts of not experienced events: the Memory Characteristics Questionnaire (Johnson, Foley, Suengas, & Raye, 1988) and the Aberdeen

*Correspondence to: Amanda J. Barnier, School of Psychology, University of New South Wales, NSW 2052, Australia. E-mail: a.barnier@unsw.edu.au

Contract/grant sponsors: University of New South Wales; Australian Research Council Queen Elizabeth II Fellowship; Australian Research Council Discovery Project; contract/grant number: DP0449447.
Contract/grant sponsor: German Science Foundation (Deutsche Forschungsgemeinschaft); contract/grant number: Sp262/3-2.

Copyright © 2005 John Wiley & Sons, Ltd.

Report Judgment Scales (ARJS; S. L. Sporer, paper presented at the biennial meeting of the American Psychology-Law Society in Redondo Beach, California, march 1998).

## CONTENT ANALYSIS TOOLS

German forensic experts designed the first tool to analyse the content of children's statements of alleged sexual abuse: *Criteria-Based Content Analysis* (CBCA; Steller & Köhnken, 1989). CBCA is part of a three-stage procedure to evaluate these statements, along with a structured interview and a validity checklist (see Raskin & Esplin, 1991). CBCA rests on the 'Undeutsch hypothesis' that certain features are found more often in genuine accounts of experienced events than in fabricated accounts of events that have not been experienced (Raskin & Esplin, 1991; Steller & Köhnken, 1989; Undeutsch, 1982). For example, genuine accounts should contain more unusual details, more reproductions of conversation, and should be more logical than fabricated accounts. Trained raters evaluate statements for the presence and absence of the 19 criteria, but to date there is no cut-off score at which an account is labelled genuine or fabricated (Pezdek et al., 2004; Zaparniuk, Yuille, & Taylor, 1995). Although originally designed for children's statements of alleged sexual abuse, CBCA has been used for adults' and children's statements, regardless of the topic (e.g. Köhnken, Schimossek, Aschermann, & Höfer, 1995; Vrij, Akehurst, Soukara, & Bull, 2002; Zaparniuk et al., 1995).

CBCA's ability to distinguish between adults' genuine and fabricated accounts ranges from a low of 47% (invented accounts scored by untrained raters; Landry & Brigham, 1992) to 93% (deceptive accounts scored by trained raters; Köhnken et al., 1995). Overall, in a review of 12 CBCA studies, Vrij (2000) concluded that the use of CBCA led raters to correctly classify 76% of genuine statements and 70% of fabricated statements. These results suggest that trained raters are able to discriminate between adults' genuine and fabricated accounts, typically at rates better than chance (Köhnken et al., 1995; Vrij, Edward, Roberts, & Bull, 2000). However, there is a lack of theory underpinning CBCA and researchers have turned to the source monitoring framework (SMF; Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2000) to provide a theoretical account of the qualitative differences between genuine and fabricated statements (Davies, 2001; Porter & Yuille, 1996; Sporer, 1997, 2004; Vrij, 2000; Vrij et al., 2000).

According to the SMF, people decide the origin of their memories by evaluating the qualitative characteristics and using a reasoning process. Memories that contain a lot of sensory and perceptual detail are attributed to an external experience, whereas memories that contain a lot of cognitive information are attributed to an internal experience, such as a dream or thought. In everyday life, people might decide whether they dreamed or thought about a particular event (internal source monitoring); whether a particular person told a funny joke (external source monitoring), or whether they only imagined something or it really happened (reality monitoring; Johnson & Raye, 1981).

People's source monitoring decisions—and in particular, their reality monitoring decisions—are not always accurate. The greater the similarity between memories from two different sources, such as genuine memories and fabricated memories, the greater the chance people will make an error. For example, when adults imagine fictitious childhood events, during which they generate sensory and perceptual details, they can make reality monitoring errors and become more confident that these events really happened (Garry, Manning, Loftus, & Sherman, 1996; Sharman, Garry, & Hunt, in press). Moreover, the

Copyright © 2005 John Wiley & Sons, Ltd. 

more often adults imagine events, the more realistic the memories of the fabricated events become, and the more likely they are to be mistaken for a real experience (Goff & Roediger, 1998; Thomas & Loftus, 2002).

In a forensic context, reality monitoring is the most important type of source decision. However, the focus is not always on whether people can tell the difference between their own memories of genuine experiences and fabricated events. Instead, it may also be on whether people can tell the difference between other people's memories of genuine experiences and fabricated events; that is, *interpersonal source monitoring* (Johnson, Bush, & Mitchell, 1998). When other people's memories contain a lot of perceptual, contextual and emotional details, they are more likely to be attributed to a real experience than when they contain a lot of cognitive detail (Johnson et al., 1998; Johnson & Suengas, 1989; Keogh & Markham, 1998).

## DISCRIMINATING GENUINE VS. FABRICATED ACCOUNTS

Many researchers have used interpersonal source monitoring,[1] CBCA, or both to discriminate between genuine and fabricated accounts. Interpersonal source monitoring can be operationalized via the MCQ (Johnson et al., 1988), which assesses the qualitative characteristics of an account, such as clarity and the amount of sensory and perceptual detail.

Comparisons between CBCA and the MCQ have shown they are related, overlapping measures of the qualitative characteristics in people's statements (Sporer, 1997, 2004). For example, Sporer (1997) videotaped participants giving accounts of events they had experienced and events they had made up (the experimenter made no mention of lying to participants). Two trained raters judged the transcripts using CBCA and then a modified version of the MCQ. Using CBCA criteria, raters correctly classified 70% of genuine accounts and 60% of invented accounts. Using MCQ criteria, raters correctly classified 71% of genuine accounts and 68% of invented accounts. Interestingly, when CBCA and the MCQ ratings were combined, raters were able to accurately classify 79% of the genuine and invented accounts overall. Vrij et al. (2000) also combined CBCA and interpersonal source monitoring criteria (although not measured with the MCQ) to distinguish between nursing students' genuine and deceptive accounts of a theft. All participants gave their account in an interview, and participants who provided the deceptive accounts were given a motivation to lie. They were told that nurses might need to tell lies and that they should be as convincing as possible. Trained raters correctly classified 75% and 72% of the accounts using CBCA or interpersonal monitoring criteria respectively, and 78% of accounts using both.[2]

As a result of these studies demonstrating the benefit of combining CBCA and interpersonal source monitoring, S. L. Sporer (paper presented at the biennial meeting of the American Psychology-Law Society in Redondo Beach, California, March 1998) created the ARJS. The ARJS examine the global characteristics, quantity and precision of details, internal processes, social aspects, and autobiographical memory of a person's account (Sporer et al., 2000). They consist of 52 items that can be grouped into 13 scales, such as *realism and coherence*, *time information*, and *lack of social desirability*, which are used to discriminate between genuine and fabricated accounts.

[1]Although we use the term interpersonal source monitoring, it is frequently referred to as reality monitoring in the deception detection literature.

[2]Note that these classifications were achieved via multiple discriminant analysis, not raters' credibility judgments. Raters' classifications using these criteria are not likely to be as accurate (Sporer, 2004).

Copyright © 2005 John Wiley & Sons, Ltd. 

## THE CURRENT EXPERIMENT

One aim of our experiment was to examine whether people's genuine, imagined, and deceptive accounts of positive and negative childhood events are qualitatively different using the MCQ and the ARJS. Imagined and deceptive accounts are not always differentiated in the literature, and could lead to different judgments about whether an event was experienced or not. When people distinguish between imagined and deceptive accounts they use internal source monitoring and research has shown that people are good at discriminating between two internally generated events, such as saying a word aloud and saying a word to yourself, and writing a word or saying it (Degl'Innocenti & Baeckman, 1996; Parker, 1995; Wippich, 1995). However, these studies involved personal source monitoring—people's judgments about their own memories—rather than interpersonal source monitoring. One would expect that people could discriminate between their own imagined and deceptive accounts because their deceptive accounts should rarely be accompanied by a belief that the event really happened (unlike imagined events; e.g. Garry et al., 1996; Hyman & Pentland, 1996). Indeed, Porter, Yuille, and Lehman (1999) demonstrated that false memories of suggested, imagined events were qualitatively different than deceptive accounts. The deceptive accounts were more vivid and clear, more coherent, and richer in detail than the false memories, which suggests that individuals can qualitatively discriminate between their own memories of imagined and deceptive events. However, we do not know whether people can discriminate between other people's imagined and deceptive accounts. Thus, another aim of our experiment was to further refine the SMF by examining whether people can discriminate between two memories that were internally generated (that is, imagined or deceptive) by other people.

In our experiment, participants produced three types of accounts. The *genuine* account was a genuine statement about a self-experienced event. The *imagined* account was a fabricated statement about an event that never happened, which participants merely imagined. The *deceptive* account was also a fabricated statement about an event that never happened. However, participants wrote about the event in order to make someone else believe that it really happened. Half of our participants wrote about negative events, whereas the other half wrote about positive events. We focused on childhood events—in particular, positive and negative childhood events—in order to maximize the potential similarity between the genuine and false events (Johnson et al., 1988). This was based on Johnson et al.'s finding that individuals rated their own genuine and imagined recent experiences as qualitatively quite different (viz. on 20 of 38 MCQ items), but rated their genuine and imagined childhood experiences as very similar (viz. on 35 of 38 MCQ items). Thus, we used childhood events in our experiment to provide a strong test of interpersonal source monitoring based on account quality.

We varied the emotional valence of the events because research suggests that positive and negative events are qualitatively different and may be remembered and forgotten at different rates (Byrne, Hyman, & Scott, 2001; D'Argembeau, Comblain, & Van der Linden, 2003; Porter & Birt, 2001). For example, D'Argembeau et al.'s (2003) participants rated their positive, neutral, and negative autobiographical memories differently. Positive events contained more visual details, odours, taste, location and time details, whereas negative events contained more details about setting and storyline, and more information about negative emotions. In other research, Detsun and Kuiper's (1999) participants wrote about pleasant and stressful real and imagined events. Both real and imagined pleasant

Copyright © 2005 John Wiley & Sons, Ltd. 

events contained more visual details, smells, tastes, and more information about location, time, and setting than the real and imagined stressful events.

Based on previous research, we predicted that there would be qualitative differences between the genuine and fabricated (imagined and deceptive) accounts. We also expected qualitative differences between the imagined and deceptive accounts. Although both accounts primarily involve imagination, the instruction to deceive should affect the qualitative details of the deceptive accounts. There are two possible predictions we can make: [1] that the deceptive accounts should be more qualitatively similar to genuine accounts than the imagined accounts, or [2] that the imagined accounts should be more qualitatively similar to the genuine accounts than the deceptive accounts. The first prediction comes from the deception and personal source monitoring literatures. When writing the deceptive accounts, participants should use what they know about making a story more believable as well as the criteria for judging their own memories as true. For instance, Vrij (2000) showed that laypeople believe that liars give fewer plausible answers, direct answers, and self-references, and give more irrelevant information and overgeneralized statements. Based on this knowledge, our participants might actively strive to be convincing by making their deceptive accounts more like their genuine accounts.

The second prediction comes from the imagination inflation literature. When people imagine events, they generate many perceptual and sensory details. Memories of imagined events that contain lots of qualitative detail are easily confused with memories of real experiences (Garry et al., 1996; Thomas & Loftus, 2002; see Pezdek & Taylor, 2000, for a discussion of the qualitative differences between genuine and fabricated accounts). Therefore, when our participants imagine events they have not experienced, they should generate many qualitative details, which should make their accounts of the imagined events similar to accounts of genuine events. Although our procedure was based on imagination inflation, we did not intend to investigate the qualitative differences between genuine accounts and fabricated accounts that participants imagined and subsequently believed in. Rather, consistent with a more forensic context, we were interested in the qualitative differences between accounts of genuinely experienced events and accounts of never experienced events that were only imagined or constructed with an intention to deceive. That is, we wished to equate the two false accounts in the sense that participants knew they had not experienced the events and tried to create plausible accounts.

We also predicted that raters who scored the accounts using either the MCQ or the ARJS should be influenced by their ratings when judging whether the described events were genuinely experienced. In other words, the amount of qualitative details in an account should influence raters' judgments of experience. Finally, we predicted that positive and negative events would be qualitatively different.

## METHOD

### Participants

Ninety psychology undergraduates (60 women and 30 men, $M = 19.89$ years, $SD = 4.41$) from the University of New South Wales, Sydney, Australia received course credit for their participation.

Copyright © 2005 John Wiley & Sons, Ltd. 

## Design

We used a mixed 3 (event: genuine, imagined, deceptive) x 2 (emotional valence: positive or negative) design. Event was the within subject variable and emotional valence was the between subject variable.

## Materials and procedure

*Phase 1*

Participants were tested in groups ranging in size from 1 to 10. Each participant sat at a computer and completed the Life Events Inventory (LEI; Garry et al., 1996). Two versions of the LEI were used: the positive version consisted of 17 positive events (e.g. 'You participated in a wedding'), and the negative version consisted of 17 negative events (e.g. 'You had to go to the emergency room late at night'). All events were taken from Garry et al. (1996) and Paddock et al. (1999). Participants indicated how certain they were that each event had happened to them before the age of 10 on a scale from 1 (*definitely did not happen*) to 8 (*definitely did happen*). Participants' ratings were used to select three target events: one event that definitely did happen (rated as 8) and two events that definitely did not happen (rated as 1). While the experimenter was selecting participants' target events and preparing materials for the next phase, participants completed a filler task.

Participants were presented with a sheet containing their three target events (labelled Event 1, Event 2 and Event 3) and an instruction booklet. For the *genuine* event, participants used a computer to write about the event they rated as '8' on the LEI. They were told to 'write about this event and what happened to you. Spend the next two minutes thinking about how you are going to write about this event. During this time remember the event happening to you.'

For the *imagined* event, participants wrote about one of the events they rated as '1' on the LEI. They were told to 'write about this event as if it happened to you. Spend the next two minutes thinking about how you are going to write about this event. During this time, imagine the event happening to you. Use the following questions to help you picture the event in your mind's eye: What are you doing while the event is occurring? What can you see happening? What can you hear? What are you thinking? What are you feeling?'

For the *deceptive* event, participants wrote about the other event they rated as '1' on the LEI. They were told to 'write about this event as if it happened to you. Spend the next two minutes thinking about how you are going to write about this event. We would like you to write about this event so that if someone who did not know whether this event had happened to you were to read your account, they would believe that this event had in fact happened to you.'

For each event, after the 2 min preparation time was up, participants were given 10 min to write their description. The order of the instructions was counterbalanced across participants.

*Phase 2*

The 270 accounts were coded so that the raters did not know whether the account was genuine, imagined, or deceptive. Each rater received training with the appropriate scale, and the ARJS rater also reviewed Sporer et al.'s (2000) training manual. Both raters practised on three pilot accounts from the current experiment, and three pilot accounts from S. L. Sporer (paper presented at the biennial meeting of the American Psychology-Law Society in Redondo Beach, California, March 1998).

Copyright © 2005 John Wiley & Sons, Ltd. 

Rater 1 rated all 270 accounts in random order using a 17-item modified version of the MCQ (Johnson et al., 1988). The version of the MCQ used in this experiment was based on its five-factor structure (Suengas & Johnson, 1988). These five factors measured *clarity* (clarity, vividness and amount of visual detail); *sensory details* (amount of smell, touch and sound details); *contextual details* (information about the spatial location of the setting, objects and people in the account); *thoughts and feelings*; and *intensity of feelings*. Each item in each factor was rated from 1 to 7; for example, 'The writer's memory for the event involves sound,' was rated from 1 (*none or a little*) to 7 (*a lot*). A mean scale score was calculated for each of the five scales by averaging the ratings for the questions included in that scale. Finally, Rater 1 judged whether the writer had experienced the event, from 1 (*definitely did not*) to 10 (*definitely did*), and the confidence in her judgment from 1 (*not at all confident*) to 10 (*absolutely sure*).

Rater 2 rated all 270 accounts in a random order using the 51-item ARJS (S. L. Sporer, paper presented at the biennial meeting of the American Psychology-Law Society in Redondo Beach, California, March 1998; Sporer, 2004). The ARJS assess the quality and quantity of details using 13 content domains: *realism and logical structure, clarity and vividness, details, spatial details, time details, thoughts, sensory impressions, emotions and feelings, memory processes and rehearsal, non-verbal and verbal interactions, extra-ordinary details, complications and unusual details, social undesirability*, and *personal significance*. Each item in each domain was rated on a 3, 4, or 7-point Likert scale; for example, 'The event described appeared to have been of personal significance to the person' from 1 (*little significance*) to 7 (*important*). A mean scale score was calculated for each of the 13 scales by averaging the ratings for the questions included in that scale.[3] For each account, Rater 2 judged whether the writer had experienced the event, from 1 (*definitely did not*) to 10 (*definitely did*), and their confidence in their judgment from 1 (*not at all confident*) to 10 (*absolutely sure*).

To ensure inter-rater reliability, a third rater rated half the accounts using the MCQ and the other half using the ARJS. Reliability between the two raters for the MCQ scales ranged from $r = 0.49$ to $r = 0.63$, all $ps < 0.01$. For the ARJS scales, inter-rater reliability ranged from $r = 0.51$ to $r = 0.73$, all $ps < 0.01$.[4] The ratings from the third rater were not used in the following analyses.

## RESULTS

Written accounts were between 34 and 407 words long ($M = 186$, $SD = 62$). A repeated measures analysis of variance (ANOVA) showed that there were no differences in word length for the genuine, imagined, and deceptive accounts, $F(2, 170) = 2.07$, $p = 0.13$, partial eta$^2 = 0.02$. To find out whether positive and negative genuine, imagined, and fabricated accounts were qualitatively different, we examined the MCQ and the ARJS ratings.

### MCQ ratings

Figures 1 and 2 show the effects of event type and emotional valence on the five dimensions of the MCQ. Because there was no interaction between event type and

[3]See S. L. Sporer (paper presented at the biennial meeting of the American Psychology-Law Society in Redondo Beach, California, March 1998) for an alternative procedure for deriving scale scores based on weighted means.
[4]The reliabilities for each of the 13 ARJS scales are available from the last author.

Copyright © 2005 John Wiley & Sons, Ltd. 

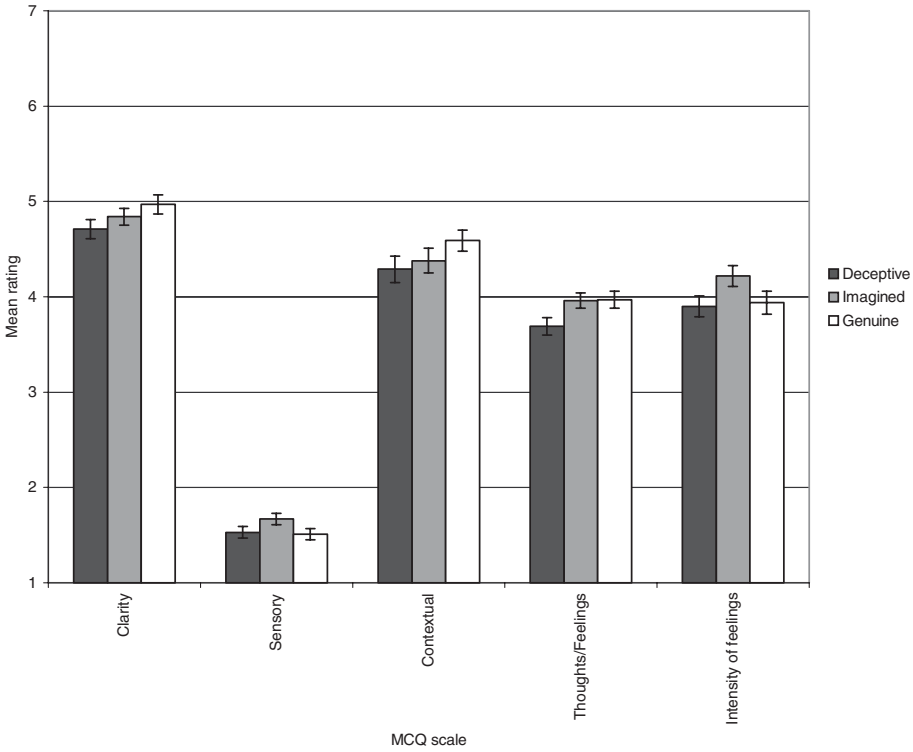


Figure 1. MCQ ratings for deceptive, imagined, and genuine events. Error bars represent the standard error of the mean

emotional valence, the results are examined separately below. Mixed 3 (event type) × 2 (valence) repeated measures ANOVAs were used to analyse the data for each scale. Follow-up pairwise comparisons were conducted where necessary.

*Event type*

Deceptive, imagined, and genuine accounts were rated differently on two factors: clarity and thoughts/feelings, $F(2, 176) = 3.47$, $p < 0.04$, partial $\text{eta}^2 = 0.04$, and $F(2, 168) = 5.07$, $p < 0.01$, partial $\text{eta}^2 = 0.05$ respectively.[5] Pairwise comparisons revealed that genuine accounts were clearer and more vivid than deceptive accounts ($p < 0.04$), and both genuine and imagined accounts contained more thoughts/feelings than deceptive accounts (genuine vs. deceptive $p < 0.05$; imagined vs. deceptive $p < 0.03$). Event type also had a marginally significant effect on the sensory details contained in the accounts, $F(2, 176) = 2.77$, $p < 0.07$, partial $\text{eta}^2 = 0.03$. Imagined accounts contained more sensory details than genuine accounts, $F(1, 88) = 4.60$, $p < 0.05$, partial $\text{eta}^2 = 0.05$.

*Emotional valence*

Figure 2 shows the effect of emotional valence on MCQ ratings. Positive and negative events were rated differently for sensory details, contextual details, and thoughts and

[5]The degrees of freedom associated with this test have been rounded to the nearest whole figure. Mauchly's test of sphericity was significant (Mauchly's $W = 0.919$, $p = 0.03$), and a Huynh-Feldt correction was used. The unrounded degrees of freedom are 1.888 and 168.050.

Copyright © 2005 John Wiley & Sons, Ltd. 

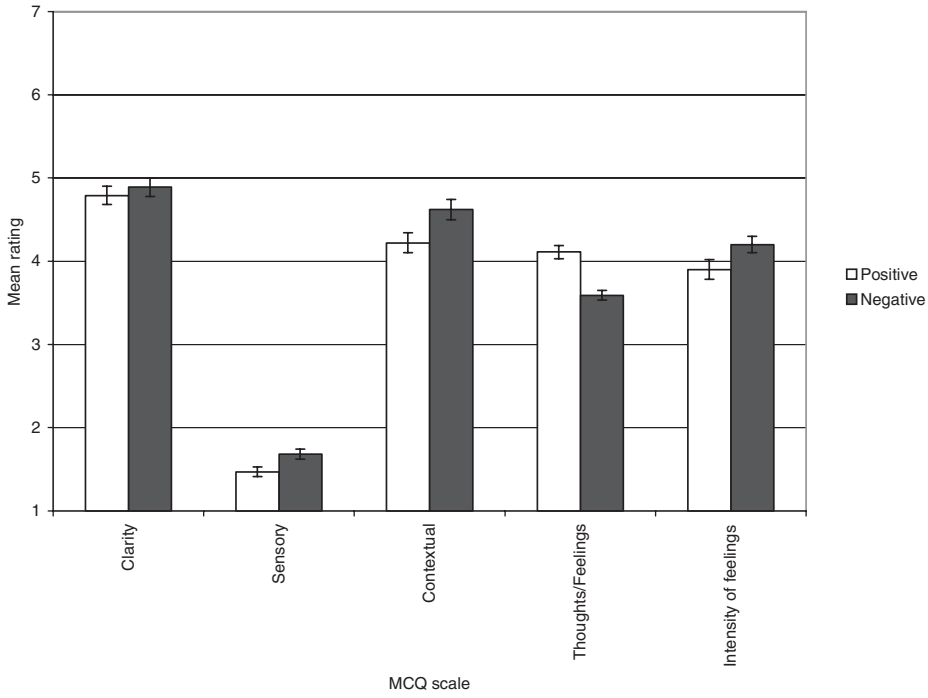


Figure 2. MCQ ratings for positive and negative events. Error bars represent the standard error of the mean

feelings. Negative accounts contained more sensory details and more contextual details than positive accounts, $F(1, 88) = 6.15$, $p < 0.02$, partial eta$^2 = 0.07$ and $F(1, 88) = 5.35$, $p < 0.03$, partial eta$^2 = 0.06$ respectively. However, positive events contained more thoughts and feelings than negative events, $F(1, 88) = 27.85$, $p < 0.01$, partial eta$^2 = 0.24$. Emotional valence also had a marginally significant effect on the rated intensity of feelings: negative events contained more intense feelings than positive events, $F(1, 88) = 3.55$, $p < 0.07$, partial eta$^2 = 0.04$.

## ARJS ratings

Figures 3 and 4 show the effects of event type and emotional valence on the 13 dimensions of the ARJS. Again there was no interaction between event type and emotional valence, so the results are examined separately below.

*Event type*

There were differences in ratings between the three types of account for time, sensory impressions, emotions, thoughts, and memory processes, all $F$s$(2, 178) > 3.07$, all $p$s $< 0.05$, partial eta$^2 = 0.03$–$0.11$. Pairwise comparisons revealed that genuine accounts contained more time details and memory processes than imagined accounts (both $p$s $< 0.01$), and more memory processes than deceptive accounts ($p < 0.05$). Imagined accounts contained more emotions and thoughts than deceptive accounts (both $p$s $< 0.02$), and more sensory impressions than genuine accounts ($p < 0.01$). Deceptive accounts contained more memory processes than imagined accounts ($p < 0.02$). Event type had a marginally significant effect on the number of core details reported, $F(2, 176) = 2.71$,

Copyright © 2005 John Wiley & Sons, Ltd. 

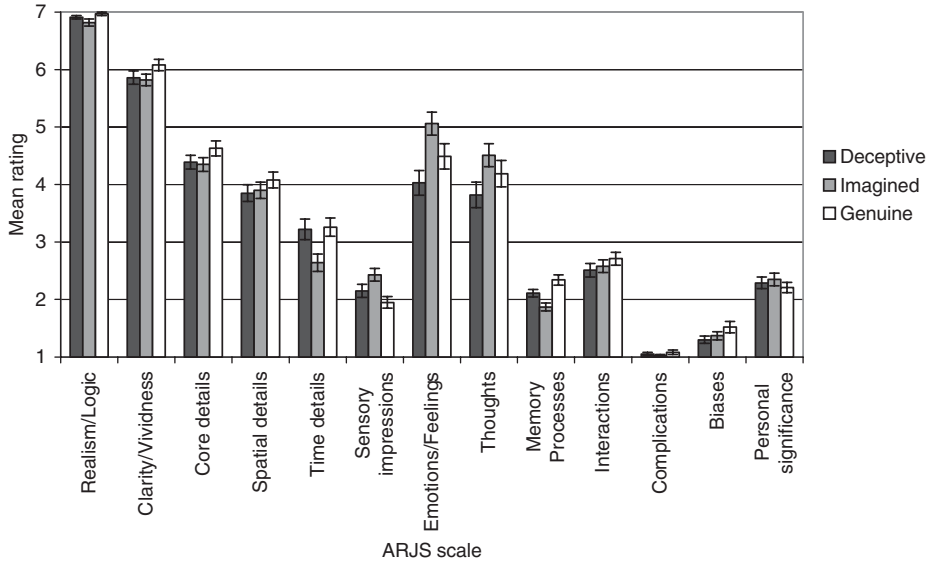


Figure 3. ARJS ratings for deceptive, imagined, and genuine events. Error bars represent the standard error of the mean

$p < 0.07$, partial eta$^2 = 0.03$. Genuine events contained more core details than imagined events, $p < 0.05$.

*Emotional valence*

Figure 4 shows the effect of emotional valence on the ARJS ratings. Negative accounts contained more spatial details, thoughts, and personal significance than positive accounts, all $F$s$(1, 88) > 3.03$, all $p$s $< 0.05$, partial eta$^2 = 0.05$–$0.40$.

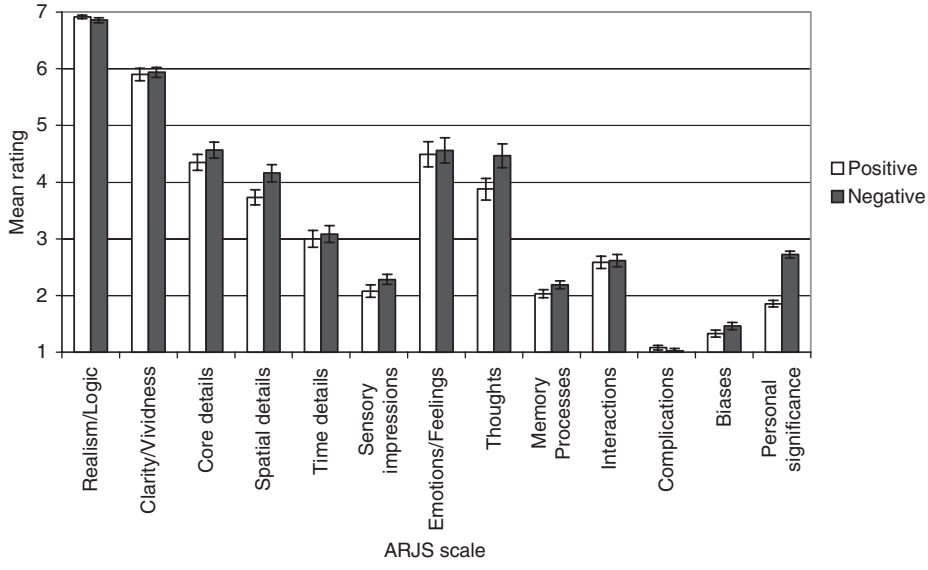


Figure 4. ARJS ratings for positive and negative events. Error bars represent the standard error of the mean

Copyright © 2005 John Wiley & Sons, Ltd. 

## Summary

Taken together, both the MCQ and the ARJS ratings showed that there were qualitative differences between genuine, deceptive, and imagined accounts, and between positive and negative accounts. Genuine accounts contained clearer details, more core details, more emotions, and thoughts and feelings than deceptive accounts, and more information about time and memory processes than imagined accounts. Imagined accounts contained more thoughts and feelings than deceptive accounts. Also, negative accounts contained more information—such as thoughts and feelings, spatial details, memory processes, and personal significance—than positive accounts.

Given that both the MCQ and ARJS revealed significant qualitative differences between genuine, imagined, and deceptive accounts, we now ask whether these differences influenced the accuracy of our trained raters' judgments of whether the described events really happened.

## Classification accuracy

We examined the accuracy of the MCQ and ARJS in discriminating between genuine and false accounts using two different methods. For each scale we [1] determined whether it was useful in discriminating genuine from false (imagined or deceptive) accounts, and [2] determined whether the rater's ratings of experience were accurate.

### *MCQ*

To determine whether the MCQ scales were useful in discriminating genuine from either type of false accounts (deceptive or imagined) we created a combined 'objective truth status' variable by coding genuine accounts as 1 and both types of false accounts as 0. A multiple discriminant analysis with the 5 MCQ criteria as predictor variables and truth status (genuine vs. imagined/deceptive) as the classifying variable revealed a nonsignificant discriminant function, Wilks' lambda $= 0.97$, $\text{chi}^2(5, N = 270) = 8.60$, ns. Overall, only 58.9% of the cases were classified correctly (64.4% of the genuine accounts and 56.1% of the false accounts).

The results of this multiple discriminant analysis can be compared with the rater's experience judgments from which one can also derive classification accuracy. We compared the objective truth status (0 vs. 1) with the 'subjective truth status' by assigning experience judgments with values 1 to 5 a value of 0, and experience judgments with values 6 to 10 a value of 1. If objective and subjective values correspond, the classification is accurate.

The MCQ rater achieved an overall accuracy rate of 65.9%. A $2 \times 3$ ANOVA revealed no reliable differences as a function of event type, $F(2, 176) = 1.72$, $p = 0.18$, partial $\text{eta}^2 = 0.02$, although classfication rates for deceptive accounts ($M = 73.3\%$) tended to be higher than for imagined ($M = 64.4\%$) and genuine accounts ($M = 60.0\%$). There was also no difference as a function of valence (positive: $M = 66.7\%$; negative: $M = 65.2\%$), or their interaction, both $F\text{s} < 1$.

The MCQ rater's confidence in her judgments did not differ for all types of account.

### *ARJS*

We performed analogous classification accuracy analyses with the ARJS. We used a multiple discriminant analysis with the 13 ARJS criteria as predictor variables and truth

Copyright © 2005 John Wiley & Sons, Ltd. 

status (genuine vs. imagined/deceptive) as classifying variable. This analysis resulted in a significant multivariate discriminant function, Wilks' lambda $= 0.90$, chi$^2(13, N = 270) = 28.56$, $p = 0.01$. Overall, 64.1% of all 270 cases were classified correctly: 58.9% of the genuine accounts and 66.7% of the imagined or deceptive accounts.

The ARJS rater's accuracy was overall 62.2% (52.2% for the genuine accounts, and 67.8% for the false accounts). More specifically, the rater classified 73.3% of the deceptive accounts, and 61.1% of the imagined accounts correctly. A $2 \times 3$ ANOVA revealed that positive accounts ($M = 69.6\%$) were classified more often correctly than negative accounts ($M = 54.8\%$), $F(1, 88) = 8.51$, $p < 0.01$, partial eta$^2 = 0.09$. The difference in accuracy of classifications between the three types of accounts was also reliable, $F(2, 176) = 4.13$, $p = 0.02$, partial eta$^2 = 0.05$. The interaction was also significant, $F(2, 176) = 3.29$, $p = 0.04$, partial eta$^2 = 0.04$. That is, positive accounts were classified more correctly than negative accounts for both deceptive and imagined events, whereas a similar number of positive and negative accounts were classified correctly for genuine events.

The ARJS rater's confidence about her judgments was affected by event type, $F(2, 176) = 219.70$, $p < 0.01$, partial eta$^2 = 0.71$. She was more confident about her judgments of deceptive accounts ($M = 6.98$, $SD = 1.78$) than genuine accounts ($M = 3.23$, $SD = 1.20$; $p < 0.01$) or imagined accounts ($M = 3.36$, $SD = 1.25$; $p < 0.01$).

Taken together, these data suggest that both the MCQ and the ARJS influenced trained raters' judgments of whether events were really experienced. Surprisingly, the tool used to rate the accounts influenced raters' confidence about their experience judgments. The MCQ rater's confidence was the same for all types of events, whereas the ARJS rater's confidence was higher for negative events and deceptive accounts.

## DISCUSSION

Our results showed that genuine, imagined, and deceptive positive and negative accounts were qualitatively different when rated using the MCQ and the ARJS. As predicted, ARJS ratings indicated that genuine accounts had more core details and memory processes than deceptive and imagined accounts, and more time details than imagined accounts. Moreover, these differential ratings influenced the rater's judgments of whether the events described in the accounts really happened. However, the MCQ was not as sensitive to potential qualitative differences between the deceptive and imagined accounts. Although MCQ ratings indicated that genuine accounts were clearer than the deceptive accounts, and both genuine and imagined accounts contained more thoughts and feelings than the deceptive accounts, the ratings did not differentiate between genuine and imagined accounts. Even so, the MCQ ratings influenced the rater's judgments of whether the events described in the accounts really happened.

Both the MCQ and the ARJS showed that the two internally generated accounts were qualitatively different. These results are consistent with Porter et al.'s (1999) finding that false memories of suggested, imagined events were qualitatively different from deceptive accounts. Our participants' imagined accounts had more thoughts and feelings (MCQ) and more emotions (ARJS) than deceptive accounts. These results support the second prediction: imagined accounts are more like genuine accounts because the act of imagining generates many perceptual and sensory details. Indeed, the imagined accounts were rated the same as the genuine accounts on the MCQ, whereas the ARJS discriminated between the two types of accounts. This may be because of our imagination instructions:

Copyright © 2005 John Wiley & Sons, Ltd. 

participants were told to think about visual details, sounds, thoughts, and feelings. In other words, these instructions told participants to generate the exact qualitative details targeted by the MCQ, which may explain why we did not find differences between genuine and imagined accounts using this tool. In contrast, the ARJS contains additional scales—such as time details and complications—that were not included in the imagination instructions, which may explain why we did find differences between genuine and imagined accounts using this tool.

As predicted, valence of the events also affected their qualitative characteristics. Negative events rated using the ARJS had more spatial details, thoughts, memory processes, and personal significance than positive events. Negative events rated using the MCQ had more sensory and contextual details, whereas positive events contained more thoughts and feelings. One possible reason for the greater qualities in negative events may be the amount of cognitive effort participants used. Conway and Pleydell-Pearce (2000) noted that we try to forget—with varying success–memories of experiences that undermine the current self, contradict our beliefs, plans, and goals, and increase anxiety or other negative emotions. Therefore, perhaps our participants used more cognitive effort to recall the negative events compared to the positive events. It is possible that this increase in cognitive effort to recall negative memories also generated more qualitative details for the negative events. In other words, when participants had to think longer and harder to recall the negative events than the positive events, they also may have generated more sensory and perceptual details for them.

The increase in cognitive effort needed to recall negative memories may also explain why negative imagined and deceptive accounts contained more qualitative information than positive imagined and deceptive accounts. Participants may have used more cognitive effort to create these negative accounts than the positive ones because they also needed to minimize any feelings of anxiety associated with negative events (de Vries, Blando, & Walker, 1995). Once again, this increase in cognitive effort for the negative accounts also may have increased the qualitative details associated with them.

Overall the ARJS discriminated between genuine and fabricated (imagined or deceptive) accounts more accurately than the MCQ. However, considering that the ARJS criteria are considered 'truth criteria'—that is their presence is likely to be associated with genuine accounts—it is surprising that more fabricated accounts than genuine accounts were classified correctly. Similarly, with the MCQ criteria evaluations of deceptive accounts tended to be more accurate than those of imagined and truthful accounts. Interestingly, when accuracy is defined as the rater's judgments of experience, the MCQ rater was more accurate in distinguishing between the genuine, imagined, and deceptive accounts than the ARJS rater. These results suggest that the tools the raters used influenced their interpersonal source judgments. Additionally, the ARJS rater gave negative events higher experience ratings than positive events. This result probably occurred because the negative events contained more qualitative details than the positive events.

Not only did the tool raters used affect their judgments of experience, it also influenced confidence in their judgments. The rater who used the MCQ was equally confident in her judgments of experience for the positive and negative genuine, imagined, and deceptive accounts, whereas the rater who used the ARJS was more confident in her judgments of the deceptive accounts than of the genuine or imagined accounts. This may be because with many more items than the MCQ, the ARJS simply provided more scope to develop a firmer view of genuineness based on account qualities.

Copyright © 2005 John Wiley & Sons, Ltd. 

There are some limitations to our experiment. First, the genuine events participants described were probably events that they had frequently rehearsed—by talking about them to others or thinking about them. The deceptive and imagined events, however, were probably events that participants had only thought about once, while generating them in the experiment. Although the fabricated events should have been largely based on real experiences or persons, participants should have incorporated other details so that the entire event was only rehearsed once. Therefore, the genuine and fabricated events might have differed from each other in terms of memory strength, which in turn, may have affected their qualitative characteristics. This limitation often occurs in research comparing true with false autobiographical events—especially childhood events—and should be considered when interpreting our findings. The second limitation is that, although significant, our inter-rater reliability was slightly lower for some MCQ and ARJS scales than others (cf. Sporer et al., 2000). This potential measurement error might have affected the reliability of our results, and again, we should be cautious in our interpretations.

Our results have both theoretical and practical implications. Theoretically, our results are important because they help refine the SMF (Johnson et al., 1993). We found qualitative differences between imagined and deceptive accounts, which lends further support to internal source monitoring, and extends it to deceptive accounts. Additionally, our results are consistent with the literature showing that the emotional valence of an event can influence the qualitative details associated with it (Byrne et al., 2001; Detsun & Kuiper, 1999). However, more research needs to be conducted to clarify which details are reliably influenced by emotional valence, and the mechanism for this effect.

We also need more research to investigate the differences between generating imagined and deceptive accounts. We know quite a lot about how imagination might enhance the qualitative details of an account, but we do not know much about how deception works. Indeed, we focused on the qualitative differences between accounts of genuinely experienced events and accounts of events that participants knew did not happen in this experiment. It is possible that participants who imagined the not-experienced events may become more confident that these events really happened, and the impact of belief in imagined accounts (somewhat similar to internalized false confessions; see Gudjonnson, 2003) would be an interesting next step in this research. Indeed, research with false confessions shows that when people believe that they have done something, they can give a compelling account of it—even when the event never happened (see, for example, Ofshe, 1992). Similarly, if people genuinely believe that their false account is a true version of events, they may give a qualitatively different account than people who know that their false event is deceptive.

Finally, future research might investigate the effects of more extreme events (rather than the mildly positive or negative events used in our experiment), more recent events, and the strength of deception. In our experiment, participants were not strongly motivated to provide a convincing account: they typed their description into a computer and there were no consequences for not giving a convincing account. Indeed, in other research we have found that deceptive accounts can match—if not exceed—the qualitative characteristics of genuine accounts of emotional, unemotional, and traumatic events when participants are given a strong motivation to lie (A. J. Barnier, S. J. Sharman, P. Ashkar, J. Leland, A. Marsh, & K. M. McConkey, in preparation; Interpersonal source monitoring of genuine and fabricated autobiographical memories: The role of deception, emotional valence, and instruction). In this research, we increased the external validity by adapting the real/

Copyright © 2005 John Wiley & Sons, Ltd. 

simulating procedure from the hypnosis literature (Orne, 1959). Participants were told to convince an experimenter (blind to condition) of their deceptive account during a one-on-one interview. In this experiment we also found that genuine and deceptive accounts were qualitatively different: deceptive accounts were rated as more realistic and clearer than the genuine accounts by trained raters using the ARJS, which indicates that participants' motives to deceive can influence the type and amount of qualitative details they provide.

Our results are also important practically for at least three reasons. First, our results support the general view that a formal analysis of the qualities of accounts can provide some indication of their source. This is relevant to professionals in any context where interpersonal source monitoring influences professional decisions. However, it is important to note that our accounts did not differ on every dimension and that raters' experience ratings, although different across accounts, were around the midpoint. Indeed, it should be acknowledged that on many qualitative dimensions the genuine, imagined, and deceptive accounts were quite similar. This suggests that we are still not in a position to provide a profile and cut-offs for ratings that definitely indicate whether an account is genuine or fabricated. Second and relatedly, our results suggest that, because of its large number of items and its scope for detecting differences, the ARJS is more useful than the MCQ in discriminating genuine, imagined, and deceptive accounts (when the person does not have a strong motivation to lie). However, given that a strong motivation to deceive may be a particular feature of some professional contexts (e.g. forensics), more research is needed to confirm the value of the ARJS for such accounts. Third, and finally, we showed that there are qualitative differences between imagined and deceptive accounts—even though our deception manipulation was relatively mild—which means that researchers and professionals should be cautious about specifying the types of fabricated accounts they use or are dealing with, as well as generalizing from one type of account to another.

## ACKNOWLEDGEMENTS

This research was supported in part by a University Research Support Program grant from the University of New South Wales. Amanda Barnier was supported by an Australian Research Council Queen Elizabeth II Fellowship and Stefanie Sharman was supported as a Postdoctoral Fellow by an Australian Research Council Discovery Project grant (DP0449447) and the University of New South Wales. Preparation of this manuscript was also supported by a grant from the German Science Foundation (Deutsche Forschungsgemeinschaft (DFG): Sp262/3-2) to the last author. Lisa McKay is now located at the Headington Institute, California. We thank Ira Hyman and an anonymous reviewer for comments on an earlier draft of this manuscript.

## REFERENCES

Byrne, C. A., Hyman, I. E., & Scott, K. L. (2001). Comparisons of memories for traumatic events and other experiences. *Applied Cognitive Psychology, 15*, S119–S133. DOI: 10.1002/acp.837

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*, 261–288.

D'Argembeau, A., Comblain, C., & Van der Linden, M. (2003). Phenomenal characteristics of autobiographical memories for positive, negative, and neutral events. *Applied Cognitive Psychology, 17*, 281–294.

Copyright © 2005 John Wiley & Sons, Ltd. 

Davies, G. M. (2001). Is it possible to discriminate true from false memories? In G. M. Davies, & T. Dagleish (Eds.), *Recovered memories: Seeking the middle ground* (pp. 153–176). Chichester, UK: John Wiley & Sons.

de Vries, B., Blando, J. A., & Walker, L. J. (1995). An exploratory analysis of the content and structure of the life review. In B. K. Haight, & J. D. Webster (Eds.), *The art and science of reminiscing: Theory, research, methods, and applications* (pp. 123–137). Washington, DC: Taylor & Francis.

Degl'Innocenti, A., & Baeckman, L. (1996). Aging and source memory: influences of intention to remember and associations with frontal lobe tests. *Aging Neuropsychology & Cognition, 3*, 307–319.

Detsun, L. M., & Kuiper, N. A. (1999). Phenomenal characteristics associated with real and imagined events: the effects of event valence and absorption. *Applied Cognitive Psychology, 13*, 175–186.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*, 208–214.

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

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. New York, NY: John Wiley & Sons.

Hyman, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory & Language, 35*, 101–117.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67–85.

Johnson, M. K., & Suengas, A. G. (1989). Reality monitoring judgments of other people's memories. *Bulletin of the Psychonomic Society, 27*, 107–110.

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

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

Johnson, M. K., Bush, J. G., & Mitchell, K. J. (1998). Interpersonal reality monitoring: judging the sources of other people's memories. *Social Cognition, 16*, 199–224.

Kapardis, A. (1997). *Psychology and law: A critical introduction.* Cambridge: Cambridge University Press.

Keogh, L., & Markham, R. (1998). Judgments of other people's memory reports: Differences in reports as a function of imagery vividness. *Applied Cognitive Psychology, 12*, 159–171. DOI: 10.1002/(SICI)1099-0720(199804)12:2<159::AID-ACP506>3.0.CO;2-J

Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology, 80*, 671–684.

Landry, K. L., & Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior, 16*, 663–676.

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: attributing mental experiences. In E. Tulving, & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). London: Oxford University Press.

Ofshe, R. J. (1992). Inadvertent hypnosis during interrogation: false confession due to dissociative state: mis-identified multiple personality and the satanic cult hypothesis. *International Journal of Clinical & Experimental Hypnosis, 40*, 125–156.

Orne, M. T. (1959). The nature of hypnosis: artifact and essence. *Journal of Abnormal & Social Psychology, 58*, 277–299.

Paddock, J. R., Noel, M., Terranova, S., Eber, H. W., Manning, C. E., & Loftus, E. F. (1999). Imagination inflation and the perils of guided visualisation. *The Journal of Psychology, 133*, 581–595.

Parker, J. F. (1995). Age differences in source monitoring of performed and imagined actions on immediate and delayed tests. *Journal of Experimental Child Psychology, 60*, 84–101.

Pezdek, K., & Taylor, J. (2000). Discriminating between accounts of true and false events. In D. F. Bjorklund (Ed.), *False-memory creation in children and adults: Theory, research, and implications* (pp. 69–91). Mahwah, NJ: Lawrence Erlbaum.

Copyright © 2005 John Wiley & Sons, Ltd. 

Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G. S., Quas, J. A., Saywitz, K. J., Bidrose, S., Pipe, M., Rogers, M., & Brodie, L. (2004). Detecting deception in children: event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology, 89*, 119–126.

Porter, S., & Birt, A. R. (2001). Is traumatic memory special ? A comparison of traumatic memory characteristics with memory for other emotional life experiences. *Applied Cognitive Psychology, 15*, S101–S117. DOI: 10.1002/acp.766

Porter, S., & Yuille, J. C. (1996). The language of deceit: an investigation of the verbal clues in the interrogation context. *Law & Human Behavior, 20*, 443–458.

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: implications for the recovered memory debate. *Law and Human Behavior, 23*, 517–537.

Raskin, D. C., & Esplin, P. W. (1991). Statement validity assessments: interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment, 13*, 265–291.

Sharman, S. J., Garry, M., & Hunt, M. (in press). Using source cues and familiarity cues to resist imagination inflation. *Acta Psychologica*.

Sporer, S. L. (1997). The less travelled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11*, 373–397. DOI: 10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0

Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P.-A. Granhag, & L. Stromwall (Eds.), *Deception detection in forensic contexts* (pp. 64–102). Cambridge: Cambridge University Press.

Sporer, S. L., Bursch, S. E., Schreiber, N., Weiss, P. E., Hoefer, E., Sievers, K., & Köhnken, G. (2000). Detecting deception with the Aberdeen Report Judgment Scales (ARJS): Inter-rater reliability. In A. Czerederecka, T. Jaskiewicz, & J. Wojcikiewicz (Eds.), *Forensic psychology and law: Traditional questions and new ideas* (pp. 197–204). Krakow: Institute of Forensic Research Publishers.

Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). Dordrecht, The Netherlands: Springer.

Suengas, A. G., & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General, 117*, 377–389.

Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory & Cognition, 30*, 423–431.

Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 27–56). The Netherlands: Kluwer.

Vrij, A. (2000). *Detecting lies and deceit*. Chichester, England: John Wiley & Sons.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*, 239–263.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out?: The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law & Human Behavior, 26*, 261–284.

Wippich, W. (1995). Self-awareness and source monitoring. *Swiss Journal of Psychology, 54*, 131–140.

Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of genuine and false statements. *International Journal of Law and Psychiatry, 18*, 343–352.

Copyright © 2005 John Wiley & Sons, Ltd. 