

## Innocent But Proven Guilty: Eliciting Internalized False Confessions Using Doctored-Video Evidence

ROBERT A. NASH and KIMBERLEY A. WADE\*

*University of Warwick, UK*

### SUMMARY

More powerful computers and affordable digital-video equipment means that desktop-video editing is now accessible and popular. In two experiments, we investigated whether seeing fake-video evidence, or simply being told that video evidence exists, could lead people to believe they committed an act they never did. Subjects completed a computerized gambling task, and when they returned later the same day, we falsely accused them of cheating on the task. All of the subjects were told that incriminating video evidence existed, and half were also exposed to a fake video. See-video subjects were more likely to confess without resistance, and to internalize the act than told-video subjects, and see-video subjects tended to confabulate details more often than told-video subjects. We offer a metacognitive-based account of our results. Copyright © 2008 John Wiley & Sons, Ltd.

For a 1994 audience, seeing Forrest Gump shake hands with President J.F.K. was astonishing. Forrest Gump, the movie, was released 31 years after J.F.K.'s assassination so this clip appeared to rewrite history. Today, Hollywood producers still recreate the past, so too do amateur video-editing enthusiasts. With the advent of more powerful computers and increasingly affordable digital equipment, desktop-video editing has become simple to perform and popular. This raises an important question: Could fake-video evidence lead people to believe they committed an act they never did?

A number of recent studies have demonstrated that various forms of fake evidence can lead people to confess to things they did not do (Horselenberg, Merckelbach, & Josephs, 2003; Kassin & Gudjonsson, 2004; Redlich & Goodman, 2003). Kassin and Kiechel (1996), for instance, asked people to complete a computer-based reaction time task. Subjects were warned not to press the ALT key because doing so would cause the computer to crash. Of course, the computer was configured to crash regardless, and the experimenters consequently accused subjects of pressing the key. Half of the subjects were also confronted with false eyewitness testimony, and when this testimony followed a fast-paced reaction time task, all these subjects falsely confessed to hitting ALT. Moreover, 65% told another confederate that they had mistakenly pressed the key, which was argued to indicate that they believed (*internalized*) the false act had occurred. Kassin and Kiechel's results showed that bogus testimony is powerful; it can lead people to confess to a self-involving act that they never committed.

\*Correspondence to: Kimberley A. Wade, Psychology Department, University of Warwick, Coventry, CV4 7AL, UK. E-mail: k.a.wade@warwick.ac.uk

What might happen though, if we used fake-video evidence rather than fake eyewitness testimony? Would subjects find the fake-video compelling? Moreover, would subjects be more likely to internalize the false act if they were *shown* the false video than if they were merely *told* of the video's existence? This comparison between seeing incriminating evidence and being told that incriminating evidence exists is important because in the USA, interrogators can legally tell suspects that incriminating evidence exists, and they can fabricate evidence—such as fingerprints or hair samples—in order to elicit a confession (Gudjonsson, 2003; Inbau, Reid, Buckley, & Jayne, 2001; Leo, 1992; Young, 1996).

In the experiments we present here, we filmed subjects as they carried out a computerized gambling task that required them to take fake money from the 'bank' when they answered questions correctly, and to return fake money to the bank when they answered questions incorrectly. Later, subjects returned to the laboratory and the experimenter accused them of taking money when they should have returned it. Half of the subjects were told that incriminating video evidence existed (hereafter, the *told-video* group), and half were exposed to a doctored video that depicted them committing the act (hereafter, the *see-video* group). An important point to note about our procedure is that unlike Kassin and Kiechel's (1996) computer-crash method in which subjects always witness the target false event occurring (i.e. they see the computer crash), our subjects did not witness the false event. Indeed, the only purported evidence against them was the false-video evidence.

How, then, might false-video evidence influence the likelihood of internalized false confessions? One theoretical account that can guide our thinking is Mazzoni et al.'s metacognitive model (Mazzoni & Kirsch, 2002; Mazzoni, Loftus, & Kirsch, 2001). According to the model, to develop false beliefs subjects must first become confident that a false event happened to them. This process begins when subjects are confronted with a counterfactual suggestion. At first, subjects might say that the event did not occur because they have knowledge that refutes the suggestion, they have no memory of the event, or they do not have enough information to make a judgement. However, feeding subjects false information can transform their beliefs; it can cause them to think their memory is unreliable and it encourages them to turn to external sources to infer whether an event genuinely happened.

Based on the metacognitive model, we predicted that *telling* subjects that incriminating video evidence exists would challenge their beliefs about the reliability of their memory and lead some of them to confess. *Showing* subjects a doctored video, however, should both challenge their beliefs about the accuracy of their memory and provide an external source that they could use to infer the act occurred. We also know that seeing is believing: in both legal and everyday decision-making tasks people are more persuaded by visual than by verbal evidence (Kassin & Dunn, 1997; King, Dent, & Miles, 1991). Thus see-video subjects should be more likely to confess to the act and come to believe that the act really happened than told-video subjects.

There are, however, reasons to expect that viewing the doctored video would not increase the likelihood of false beliefs. In the metacognitive model, people will readily reject a suggested event if they judge that their absence of a memory is *diagnostic*, that is, if they decide that having no memory confirms that the event never happened (Mazzoni & Kirsch, 2002; see also Dodson & Schacter, 2001). In addition, research shows that memories of recent events—such as experiences that occurred within the last year—are packed with perceptual, contextual and semantic details (Johnson, Foley, Suengas, & Raye, 1988; Johnson, Hashtroudi, & Lindsay, 1993). We might predict, then, that subjects would

be skeptical if they were told they committed a mischievous act, but at the same time had no vivid memory of this event occurring. This skepticism might act as a warning; prompting subjects to systematically evaluate their memories and the reliability of the video evidence. Research shows that people can and do evaluate the source of a suggestion before accepting it as fact (Dodd & Bradshaw, 1980; Vornik, Sharman, & Garry, 2003), and people are well informed about the advances of digital software. Indeed, we asked 66 volunteers to list as many photo- and video-editing software products as they could—75% named at least one programme and 50% reported having edited digital media at some point in the past. Finally, research by Garry and Wade (2005) revealed that verbal descriptions of a false event were more likely to elicit illusory beliefs than doctored photographs. Thus, subjects might not find our doctored-video evidence compelling.

In sum, our goals were threefold: to test whether viewing doctored-video evidence would elicit more false confessions and false beliefs than merely being told that video evidence exists, to develop a new paradigm for examining the effect of fabricated evidence on false confessions, and to gather information about possible cognitive influences.

## EXPERIMENT 1

### Method

#### *Subjects*

Thirty students (13 males, 17 females;  $M = 21.20$  years,  $SD = 2.48$ , range = 18–27) at Warwick University received £6 for participating. We randomly allocated them to the *see-video* or *told-video* condition and they took part individually.

#### *Procedure*

The University of Warwick Humanities and Social Sciences Research Ethics Committee approved our procedure. Two experimenters and one confederate conducted the experimental sessions.<sup>1</sup> To ensure that the experimenters and confederate behaved consistently and similarly across subjects, they were extensively trained to follow an interview protocol. Subjects always met with the same experimenter at Sessions 1 and 2.

*Session 1.* In Session 1, subjects completed a computerized gambling task. They were told that the experimenters were investigating the effects of gambling with physical and electronic credit. All subjects were told that they were in the ‘physical credit’ condition and that they would be filmed completing the gambling task in two separate sessions. Subjects were seated in front of a computer, and given a pile of fake money to gamble with and a pile that represented the *bank*. They were told that their gambling money was the previous subject’s winnings, and the aim of the task was ‘to *increase* those winnings by as much as possible for the next subject’. In reality, subjects always began with the same amount of money (£115). Next, the experimenter gave the subject an opportunity to ask questions, started the video-camera recording in full view of the subject, and left the room. When the subject finished, the experimenter re-entered the room and reminded the subject to return for Session 2 (approximately 2–3 hours later) to complete a second gambling task.

<sup>1</sup>Our design restricts us to gathering one data point per subject on each conformity measure, thus to minimize procedural variation across subjects we used only two experimenters and one confederate in each experiment.

**Q6: Of what is "Rhytiphobia" the fear?**  
 Think carefully, then place a bet on ONE answer ONLY!

(A)	<b>Getting wrinkles (Odds = 2:1)</b>	£ <input type="text"/>	<input type="button" value="Enter a bet"/>
(B)	<b>Getting dirty (Odds = 3:1)</b>	£ <input type="text"/>	<input type="button" value="Enter a bet"/>
(C)	<b>Getting undressed (Odds = 5:1)</b>	£ <input type="text"/>	<input type="button" value="Enter a bet"/>
(D)	<b>Getting leprosy (Odds = 10:1)</b>	£ <input type="text"/>	<input type="button" value="Enter a bet"/>

Figure 1. Screen shot of a question in the gambling task

*The gambling task.* We created a general knowledge, multiple choice, computerized gambling task. There were 15 questions, and each question had four possible responses. Each response was associated with a different odds ratio (Figure 1). Subjects selected an answer and typed in the amount of money they wished to gamble on each question. We piloted the task, and adjusted the difficulty of the questions until pilot subjects performed at 33% accuracy. Subjects received feedback after responding to each question. If their response was correct, a green tick appeared with instructions to take money from the bank. If their response was incorrect, a red cross appeared with instructions to return money to the bank. The task took approximately 15 minutes to complete.

*Creating the video.* In between the two sessions, we used an iMac and Final Cut Pro 5<sup>®</sup> to create the doctored video. We took a 10- to 20-second segment that showed the subject answering a question correctly and digitally replaced the green tick on the monitor with a red cross (Figure 2). The resulting clip ostensibly showed the subject collecting money from the bank when they should have returned money.

*Session 2.* In Session 2, the experimenter informed the subject that there was a problem with the earlier session: the video showed that the subject took money from the bank



Figure 2. The video doctoring process

instead of returning it when they answered a question incorrectly. The experimenter told the subject that the incident meant her or his data were invalid, as were the data from the next subject who inherited the wrong sum of money. See-video subjects also viewed a doctored video clip that depicted them committing this act. They were allowed to watch the clip through twice if they requested.

Next, subjects were asked to sign a confession form (Appendix), that had been handwritten by the experimenter prior to the session, to confirm that they took money from the bank when they should have returned it. They were told that if they signed the confession they would *not* receive their £6 payment, and if they did not sign the confession they would have to meet with the professor in charge of the study to find out whether or not they would receive payment. If the subject refused to sign the confession, the experimenter said that the professor was likely to refuse payment because the video clearly showed that the subject took the money. The experimenter then asked the subject once more to sign the statement. These two requests to sign the confession served as our measure of compliance.

Subjects who did not sign were asked to write down what they thought the study was investigating.<sup>2</sup> This enabled us to determine whether subjects figured out the true nature of the experiment. Subjects who signed the confession were also asked to write down what they thought the study was investigating and they were asked to sit in a waiting room while the experimenter spoke with the professor in charge. A confederate—blind to the experimental hypotheses—was posing as another subject in the waiting room. She initiated a conversation according to the protocol and encouraged the subject to describe the accusation by enquiring which study she or he was participating in and why they were waiting. Unless the subject clearly stated that they did not believe they had committed the act, the confederate asked whether she or he could remember how the error occurred. This conversation was covertly recorded using a digital voice recorder and served as our measure of internalization and confabulation. The confederate was always blind to the condition to which subjects were assigned, but on some occasions this blind was broken because the subject mentioned being told about or viewing the video evidence. We return to this issue shortly.

Finally, subjects returned to the laboratory and the experimenter asked if she or he could describe how the error occurred. If the subject speculated about a possible scenario or confabulated details, the experimenter asked the subject to write the details down on the back of the confession statement. These details served as an additional measure of confabulation. Subjects were debriefed and asked to provide retrospective consent for having their conversation with the confederate recorded. All subjects gave consent.

#### *Scoring compliance, internalization and confabulation*

Recall that the confederate did not always remain blind to each subject's condition. To ensure that the confederate did not lead the subject when the blind was broken, we screened transcripts of subjects' discussions with the confederate. The confederate never deviated from the interview protocol, thus we are confident that see-video and told-video subjects were treated equally.

Subjects were judged as having complied if they signed the confession statement on the first or second request. Two trained observers reviewed transcripts of the subjects'

<sup>2</sup>In fact, all subjects in Experiment 1 signed the confession form. However, we used this procedure in Experiment 2 for subjects who did not confess.

discussions with the confederate and used Redlich and Goodman's (2003) trichotomous coding scheme to determine whether each subject internalized the act and confabulated. To ensure that the observers were blind to the see-video/told-video manipulation, we removed any comments from the transcripts that revealed which condition the subject was in. The observers categorized subjects as *no internalization*, *partial internalization* or *full internalization*. To be categorized as *partial internalization* subjects had to make statements indicating that they believed they might have committed the act ('I think I messed up his experiment'), and to be categorized as *full internalization* they had to make statements indicating that they believed they did commit the act ('I took money when I was supposed to give it back'). The observers categorized subjects as *no confabulation*, *hypothesizing* or *full confabulation*. To be categorized as *hypothesizing* subjects were required to speculate about what could have happened, without implying that was how it *did* happen ('I probably expected I was right, and didn't take any notice of the cross'), and to be categorized as *full confabulation* subjects were required to describe how the act occurred ('I was concentrating so hard on the money, I forgot to give rather than take when I was wrong'). Observers agreed on 83% ( $\kappa = .67$ ) of internalization categorizations and 83% ( $\kappa = .47$ ) of confabulation categorizations. When observers disagreed, the more conservative categorization was accepted. The video recordings showed that none of the subjects took money from the bank when they were not supposed to.

## Results and discussion

Most subjects seemed to be extremely surprised when the experimenter revealed the true nature of the study. For example, one subject replied 'You're kidding? I really thought I did that'! Moreover, subjects' notes about the purpose of the experiment showed that all of them believed the study aimed to investigate gambling behaviour and not false confessions or any topic related to memory distortions, compliance and so on. Together these findings suggest that our results are not the product of experimental demand.

### *Rates of compliance, internalization and confabulation*

All subjects signed the confession form (complied) either on the first or second request. This finding replicates the extraordinarily high compliance rates found in previous false evidence studies (Kassin & Kiechel, 1996; Redlich & Goodman, 2003). Eighty-seven per cent of subjects signed on the first request and 13% on the second request. Moreover, the observers' classifications revealed that 20% partially internalized the act and 63% fully internalized the act. Seven per cent of subjects hypothesized about why they took the money and 3% fully confabulated details about how it occurred. Taken together, these data provide further evidence that a combination of social demand, phoney evidence and false suggestion from a credible source can lead a substantial number of people to falsely confess and believe they committed an act they never did.

Recall that our primary question was whether see-video subjects were more likely than told-video subjects to confess to and believe they committed the act. To address this question we calculated the percentage of told-video and see-video subjects classified as complying, internalizing and confabulating and conducted three  $2 \times 3$  Fisher's exact tests. These data and the corresponding *p*-values are presented in Table 1. As these results show, see-video and told-video subjects were equally likely to confess, but there was a tendency for see-video subjects to confess on the first request more often than told-video subjects (row 1 in Table 1). In terms of internalization, there were significant group differences

Table 1. Experiment 1: percentage of subjects classified into each compliance, internalization and confabulation category as a function of condition (raw frequencies in parentheses;  $n = 15$  in each condition)

Condition		Categorization		Fisher's exact $p$ (two-sided)
Compliance	No compliance	Complied on second request	Complied on first request	$p = .10$
Told-video	0% (0)	27% (4)	73% (11)	
See-video	0% (0)	0% (0)	100% (15)	
Internalization	No internalization	Partial internalization	Full internalization	$p = .03$
Told-video	33% (5)	7% (1)	60% (9)	
See-video	0% (0)	33% (5)	67% (10)	
Confabulation	No confabulation	Hypothesizing	Full confabulation	$p = 1.00$
Told-video	93% (14)	7% (1)	0% (0)	
See-video	87% (13)	7% (1)	7% (1)	

(row 2 in Table 1). Taking a closer look at the internalization data, using  $2 \times 2$  Fisher's exact tests, we found that see-video subjects were no more likely to fully internalize the accusation than told-video subjects,  $p = 1.00$ . However, when we collapsed across the *partial* and *full* internalization categories we found that see-video subjects were more likely to partially or fully internalize the act than told-video subjects,  $p = .04$ . Finally, in terms of hypothesizing and confabulating (row 3 in Table 1), the two groups did not differ significantly.

These results lead us to conclude that the *doctored-video* paradigm is a powerful tool for eliciting and examining false confessions. Perhaps more importantly, our findings suggest that viewing fabricated evidence might promote internalization more than being falsely told that evidence exists. Our findings fit with the metacognitive account of false beliefs (Mazzoni & Kirsch, 2002) and warrant concern about the use of fabricated evidence in interrogations (Gudjonsson, 2003).

## EXPERIMENT 2

The false act used in Experiment 1 was designed to be more memorable than the false acts used in most confession studies, and more ecologically valid in that it involves some amount of intention, a negative consequence, and could be perceived as having a motive (but see Russano, Meissner, Narchet, & Kassin, 2005). It is possible, however, that our confession rates were high in Experiment 1 because the cheating event was still somewhat plausible. Indeed, subjects answered, on average, 10 out of 15 questions incorrectly on the gambling task, thus many subjects may have been unsure whether they were innocent or guilty when the experimenter accused them of taking money from the bank. To ensure that our findings were not the result of using a highly plausible 'crime', in Experiment 2 we attempted to replicate our findings, but we accused subjects of something much less plausible: inappropriately taking money from the bank on three occasions. To confirm that taking money on three occasions would be a less plausible accusation, we described the

experiment to 20 volunteers. We asked them how plausible it would be on a scale from 1 (very implausible) to 7 (very plausible), for them to have improperly taken money on either one or three occasions, without remembering doing so. The volunteers determined that making three mistakes was significantly less plausible ( $M = 1.75$ ) than making one mistake ( $M = 3.30$ ),  $t(19) = 5.82$ ,  $p < .0001$ ; also 80% rated three mistakes as less plausible than one mistake and 50% rated three mistakes as 'very implausible'. As in Experiment 1, subjects in Experiment 2 always sat directly in front of the screen, and the on-screen feedback was clear and large; they also knew they had made these mistakes on only a small subset of questions. Thus subjects could not feasibly conclude that three mistakes were equally as plausible as one mistake by reasoning that the feedback was ambiguous or that they could have misunderstood the instructions.

## Method

### *Subjects*

Thirty students (13 males, 17 females;  $M = 20.43$  years,  $SD = 1.94$ , range = 18–24) at Warwick University received £6 for participating. We randomly allocated them to the *see-video* or *told-video* condition.

### *Procedure*

The procedure was identical to Experiment 1, except for three modifications. First, in Session 2, instead of accusing subjects of taking money from the bank when they answered one question incorrectly, the experimenter accused subjects of taking money from the bank on three separate occasions, when they answered three questions incorrectly. Second, we doctored three sections of *see-video* subjects' original recordings so that the fake videos depicted them improperly taking money from the bank three times. To do this we needed to ensure that subjects answered at least three questions correctly. Thus we increased the length of the quiz from 15 to 18 questions, gave subjects more money to gamble with (£265) and made the questions slightly easier.

Two independent observers, blind to which condition each subject was in, categorized subjects' discussions with the confederate. The observers agreed on 80% ( $\kappa = .68$ ) of the internalization categorizations and on 93% ( $\kappa = .80$ ) of the confabulation categorizations. Once again, the more conservative of the observers' categorizations were accepted, and none of the subjects took money from the bank when they were not supposed to.

## Results and discussion

Even with a less plausible accusation, the vast majority of subjects signed the confession and internalized the false act.<sup>3</sup> Overall, 93% of subjects complied: 87% on the first request and 7% on the second request. Thirty per cent of subjects partially internalized the act and 43% fully internalized it. Ten per cent of subjects hypothesized about how it might have happened and 7% confabulated details. Except for the full internalization rate of *told-video* subjects, which was notably lower than in the previous experiment, these confession and internalization rates are similar to those in Experiment 1. Thus, even when subjects were

<sup>3</sup>Four subjects in Experiment 2 indicated in their notes that they were suspicious about the experiment. Two of these had seen the video evidence; however, neither suggested that the video might have been manipulated. None of these four subjects were included in the analyses, and four further subjects were tested. Our results are therefore based upon 30 non-suspicious subjects.

Table 2. Experiment 2: percentage of subjects classified into each compliance, internalization and confabulation category as a function of condition (raw frequencies in parentheses;  $n = 15$  in each condition)

Condition	Categorization			Fisher's exact $p$ (two-sided)
Compliance	No compliance	Complied on second request	Complied on first request	$p = .73$
Told-video	7% (1)	13% (2)	80% (12)	
See-video	7% (1)	0% (0)	93% (14)	
Internalization	No internalization	Partial internalization	Full internalization	$p = .005$
Told-video	40% (6)	47% (7)	13% (2)	
See-video	13% (2)	13% (2)	73% (11)	
Confabulation	No confabulation	Hypothesizing	Full confabulation	$p = .04$
Told-video	100% (15)	0% (0)	0% (0)	
See-video	67% (10)	20% (3)	13% (2)	

accused of committing a relatively implausible act, we obtained extremely high levels of false confessions and false beliefs.

Table 2 shows the percentage of told-video and see-video subjects classified as complying, internalizing and confabulating, and the corresponding  $2 \times 3$  Fisher's exact test  $p$ -values. As in Experiment 1, see-video and told-video subjects were equally likely to confess (row 1 in Table 2). In terms of internalization, again there were significant group differences (row 2 in Table 2). See-video subjects were more likely to fully internalize guilt than told-video subjects,  $p = .003$ . However, when we collapsed across the *partial* and *full* internalization categories, we found no significant differences,  $p = .22$ . As for the confabulation measure, in Experiment 1 we found no group differences, but in Experiment 2, see-video subjects were more likely to hypothesize *or* confabulate than told-video subjects,  $p = .04$ .

Together, these data replicate those in Experiment 1 by showing that viewing the video led to higher internalization rates than being told that the video existed. The finding that see-video subjects were more likely to hypothesize *or* confabulate than told-video subjects fits with Kassin and Kiechel's (1996) seminal computer-crash experiment which showed subjects were more likely to fabricate information if eyewitness evidence existed than if it did not.

The findings in Experiment 2 rule out a possible counterexplanation for the high confession rates observed in Experiment 1; that is, that subjects signed the confession statement because they thought the target false event was likely to have occurred. Even when subjects were accused of cheating three times—an act that independent observers rated as relatively implausible—subjects still confessed at extremely high rates. Thus, the idea that our findings were attributable to the plausibility of the accusation cannot account for our results.

Finally, the conversations subjects had with our confederate illustrate the conviction with which subjects held their false beliefs. Table 3 provides some sample conversations which demonstrate that subjects did not merely comply with the suggestion; rather, they expressed deep concern at having taken the money and at having compromised the data.

Table 3. Excerpts from conversations that subjects had with the confederate

	Full internalization	Full confabulation
Subject X (see-video group)	Subject Y (see-video group)	Subject Z (see-video group)
<i>Subject:</i> I, I kind of took money from the bank when I got them wrong <i>Confederate:</i> Okay... <i>Subject:</i> I don't know why. But I've seen the video and it's proof so it's... <i>Confederate:</i> Oh dear! <i>Subject:</i> So yeah, I've got a bit of a problem <i>Confederate:</i> I see. Do you remember that happening, or...? <i>Subject:</i> No, but I've looked at the video, and it's kind of quite clear <i>Confederate:</i> Mm... <i>Subject:</i> It's scary seeing yourself on video making a complete idiot of yourself!	<i>Subject:</i> ... it's never happened before—trust me, it's really bad! I'm just taking money, I know it's only fake money, but I'm just taking money, when it's not even mine!	<i>Subject:</i> I wasn't concentrating and obviously it's close, you know, the bank and [your own money]. I got confused

## GENERAL DISCUSSION

Ours is the first study to demonstrate the dangers of modern digital manipulation technology when encouraging people to remember self-involving, recently occurring experiences (see also Garry & Wade, 2005; Wade, Garry, Read, & Lindsay, 2002); and on a broader level our results show that seeing fake evidence is more convincing than being merely told of its existence. When our subjects viewed fake-video evidence, nearly 100% falsely confessed, and 67% (Experiment 1) and 73% (Experiment 2) believed they committed a false act. When subjects were simply told that video evidence existed, again, nearly 100% falsely confessed, and 60% (Experiment 1) and 13% (Experiment 2) developed false beliefs. The fact that our internalization ratings were derived from subjects' conversations with a confederate—who ostensibly had no involvement with the study—means that our internalization data cannot be explained by the possibility that subjects were simply conforming to experimental demand.

Legal scholars are increasingly aware that digital trickeries could be used in forensic contexts to mislead people (House of Lords Select Committee on Science and Technology, 1998; Nelson, 1997). While our cheating event differs dramatically from the crimes that suspects are accused of, in particular because the act of taking money could have been considered unintentional, our findings are proof that many people will readily confess and develop erroneous beliefs if they are accused of an act and told about or confronted with false-video evidence [see Horselenberg et al., 2006 (Study 3), and Kassin & Kiechel, 1996 for brief discussions of unintentional 'crimes']. Aside from the legal applications, our findings are also relevant to clinical settings. Kehle, Bray, Margiano, Theodore, and Zhou (2002) demonstrated that doctored videos could be used to reduce the disruptive behaviour of children with severe emotional disturbance. They showed some children videos of

themselves with all episodes of disruptive behaviour edited out—leaving only evidence of positive behaviour in the video. The children's disruptive behaviour decreased significantly after viewing the edited videos. Kehle et al. hypothesized that their successful behaviour modification technique is effective because the doctored media alters the children's beliefs, and perhaps even their memories, about how they normally behave. Our results provide preliminary support for this hypothesis.

On a theoretical level, our findings extend the metacognitive account of false belief development (Mazzoni & Kirsch, 2002). When subjects were confronted with an accusation of cheating, they failed to remember improperly taking money. Thus they turned to the available external information—the video—to help them decide whether or not they committed the act. We might hypothesize, then, that viewing the fake video made the false suggestion more powerful by one of three mechanisms: [1] increasing the perceived plausibility of the suggestion, [2] increasing the perceived credibility of the experimenter or [3] enabling subjects to speculate about how or why they took money from the bank. One point to note is that we asked subjects to write down why they took the money, if they could conceive of a possible reason. There were comments such as 'I must have been expecting to get the question right', and these comments were more common in the see-video than told-video condition. Thus, it is feasible that watching the video enabled subjects to speculate about why the false event occurred.

Models of false memory development propose that once subjects form false beliefs (Mazzoni et al., 2001) or relevant and plausible imagery (Hyman & Kleinknecht, 1999), they are only one step away from producing false memories. The final step involves misattributing internally generated details to genuine experience (as per the source-monitoring framework of Johnson et al., 1993; Lindsay, 2008; see also Henkel & Coffman, 2004). Several recent studies have shown that repeated acts or extended periods of imagination can increase the likelihood of source misattributions (e.g. Mazzoni & Memon, 2003; Read & Lindsay, 2000; Thomas & Loftus, 2002). For example, when people are asked to imagine a counterfactual experience, they are more likely to falsely remember that the event really happened if they imagined it three or five times rather than just once (Goff & Roediger, 1998). Due to ethical constraints we were unable to let our subjects consider the cheating event for an extended period, but the imagination research leads us to suspect that, given more time and the instruction to imagine, our see-video subjects might have been particularly vulnerable to memory distortions.

Our results show that the vast majority subjects who viewed the fake video confessed to the false act without resistance, and believed, at least partially, that they had genuinely taken money. Even the told-video subjects, who did not see the fake-video evidence, indicated high rates of false beliefs. Although the fake-video evidence might have produced these high confession rates, without a control condition in which subjects were accused without any evidence of taking the money we cannot say for sure whether the video evidence or the gambling-task *per se* produced the high false confession rates. Previous false evidence research (e.g. Kassir & Kiechel, 1996) indicates the false suggestion of evidence enhances internalization beyond an accusation alone. In line with this finding, many of our told-video subjects claimed they would not have believed the accusation had there been no mention of the video evidence. What we can say about our data is that subjects found the fake-video evidence compelling, and this type of corroboration can cultivate internalized false confessions.

Our doctored-video paradigm offers a novel way to investigate the influence of false evidence upon false confessions and the formation of false beliefs. Future research should

examine the combined impact of internal (metacognitive) and external (video) evidence on source-monitoring decisions. Might priming subjects' knowledge about digital manipulations cause them to favour what they know over what they see?<sup>4</sup> Moreover, there are several differences between our study and prior studies using false incriminating evidence, so we do not know whether false confessions and beliefs are more likely to emerge when the fake evidence is in the form of a video, photograph, trace evidence or eyewitness testimony. On a methodological note, our procedure for assessing internalization followed that of Kassin and Kiechel's (1996) method which relies upon linguistic differences between subjects' conversations with a confederate. Alternative procedures for assessing internalization that rely upon more overt behaviours should be explored.

Regardless of the mechanisms involved in creating our see-video *versus* told-video effect, the results show that doctored videos, or simply the proposition that video evidence exists, are potent forms of suggestion that can contribute to false confessions and foster false beliefs. According to these results, our advice to those who receive digital footage of themselves is: be warned, digital images from untrustworthy sources are like a box of chocolates; never know what you are going to get.

### ACKNOWLEDGEMENTS

We thank Derrick Watson, Stefanie Sharman, Lauren French, Linda Henkel, Alan Scoboria and two anonymous reviewers for their helpful comments on earlier drafts. We also thank Cathy Brown and Olwen Bryer for their assistance in data collection, and Rachel Wever and Jeffrey Sun for scoring the transcripts.

### REFERENCES

- Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, *19*, 695–704. DOI: 10.1016/S0022-5371(80)90379-5
- Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161.
- Garry, M., & Wade, K. A. (2005). Actually, a picture is worth less than 45 words: Narratives produce more false memories than photographs. *Psychonomic Bulletin & Review*, *12*, 359–366.
- 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*. West Sussex, England: Wiley.
- Henkel, L. A., & Coffman, K. J. (2004). Memory distortions in coerced false confessions: A source monitoring framework analysis. *Applied Cognitive Psychology*, *18*, 567–588. DOI: 10.1002/acp.1026
- Horselenberg, R., Merckelbach, H., & Josephs, S. (2003). Individual differences and false confessions: A conceptual replication of Kassin and Kiechel (1996). *Psychology, Crime and Law*, *9*, 1–8.
- Horselenberg, R., Merckelbach, H., Smeets, T., Franssens, D., Peters, G.-J. Y., & Zeles, G. (2006). False confessions in the lab: Do plausibility and consequences matter? *Psychology, Crime and Law*, *12*, 61–75.
- House of Lords Select Committee on Science and Technology. (1998). *Digital images as evidence*. London: The Stationery Office.

<sup>4</sup>We thank Alan Scoboria for this suggestion.

- Hyman, I. E., Jr., & Kleinknecht, E. E. (1999). False childhood memories. In L. M. Williams, & V. L. Banyard (Eds.), *Trauma and memory* (pp. 178–188). Thousand Oaks, CA: Sage.
- Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, MD: Aspen.
- 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*, 371–376. DOI: 10.1037/0096-3445.117.4.371
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. DOI: 10.1037/0033-2909.114.1.3
- Kassin, S. M., & Dunn, M. A. (1997). Computer-animated displays and the jury: Facilitative and prejudicial effects. *Law and Human Behavior*, *21*, 269–281. DOI: 10.1023/A:1024838715221
- Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest*, *5*, 33–67. DOI: 10.1111/j.1529-1006.2004.00016.x
- Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science*, *7*, 125–128. DOI: 10.1111/j.1467-9280.1996.tb00344.x
- Kehle, T. J., Bray, M. A., Margiano, S. G., Theodore, L. A., & Zhou, Z. (2002). Self-modeling as an effective intervention for students with serious emotional disturbance: Are we modifying children's memories? *Psychology in the Schools*, *39*, 203–207. DOI: 10.1002/pits.10031
- King, W. C., Dent, M. M., & Miles, E. W. (1991). The persuasive effect of graphics in computer-mediated communication. *Computers in Human Behavior*, *7*, 269–279. DOI: 10.1016/0747-5632%2891%2990015-S
- Leo, R. A. (1992). From coercion to deception: The changing nature of police interrogation in America. *Crime, Law and Social Change*, *18*, 35–59.
- Lindsay, D. S. (2008). Source monitoring. In H. L. Roediger III (Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference* (J. Byrne, Ed.) (vol 4, pp. 325–348). Oxford: Elsevier.
- Mazzoni, G., & Kirsch, I. (2002). Autobiographical memories and beliefs: A preliminary metacognitive model. In T. J. Perfect, & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 121–146). Cambridge, UK: Cambridge University Press.
- Mazzoni, G., & Memon, A. (2003). Imagination can create false autobiographical memories. *Psychological Science*, *14*, 186–188. DOI: 10.1046/j.1432-1327.2000.01821.x
- Mazzoni, G. A. L., Loftus, E. F., & Kirsch, I. (2001). Changing beliefs about implausible autobiographical events: A little plausibility goes a long way. *Journal of Experimental Psychology: Applied*, *7*, 51–59. DOI: 10.1037//1076-898X.7.1.51
- Nelson, E. (7 February 1997). Claims of photo fakery get lots of exposure in court. *Wall Street Journal (Eastern Edition)*, B2.
- Read, J. D., & Lindsay, D. S. (2000). “Amnesia” for summer camps and high school graduation: Memory work increases reports of prior periods of remembering less. *Journal of Traumatic Stress*, *13*, 129–147. DOI: 10.1023/A:1007781100204
- Redlich, A. D., & Goodman, G. S. (2003). Taking responsibility for an act not committed: The influence of age and suggestibility. *Law and Human Behavior*, *27*, 141–156. DOI: 10.1023/A:1022543012851
- Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M. (2005). Investigating true and false confessions within a novel experimental paradigm. *Psychological Science*, *16*, 481–486.
- Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory & Cognition*, *30*, 423–431.
- Vornik, L. A., Sharman, S. J., & Garry, M. (2003). The power of the spoken word: Sociolinguistic cues influence the misinformation effect. *Memory*, *11*, 101–109.
- Wade, K. A., Garry, M., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin & Review*, *9*, 597–603.
- Young, D. (1996). Unnecessary evil: Police lying in interrogations. *Connecticut Law Review*, *28*, 425–477.

## **APPENDIX**

### **Text of confession**

### **Reason for refusal of payment/credit**

During the first session, participant took credit from the 'bank' [in Experiment 2 add 'on three separate occasions'] when clearly instructed to return it. Due to nature of experiment, session data cannot be used for this reason. Data from subsequent session also invalid due to incorrect recording of results. Refusal requested to compensate time and money loss.