Supplementary Information for

Seeing is Believing: The Continued Influence of Known AI-Generated 'Deepfake' Videos

Simon Clark and Stephan Lewandowsky

School of Psychological Science, University of Bristol, UK

Extended Method

Video Transcripts

Experiments 1 and 2

This one-minute video appears to be a clip from a longer video, starting and stopping abruptly. It features John Carter, a fictional local politician, sitting at a table in a bar. He is wearing a suit and open shirt, and holding a glass of wine. He is facing the camera, and a half empty pint of beer on the table suggests he is speaking to someone opposite, although no one else is visible in the video. He speaks quietly but can be heard fairly clearly, in spite of the background noise of a bar. At one point he stops speaking as someone passes the table, suggesting he does not wish to be overheard. The video looks as if it could have been secretly filmed by the person opposite; the angle and lighting are not perfect but John Carter's face is clearly visible.

"… So yeah, obviously this is between you and me, OK, I haven't told anyone else this, not even the wife. … But yeah, I did it … yeah I took the money. … But to be fair, right, I'm pretty sure everyone was doing it, or at least I definitely wasn't the only one. … Because I mean, what's the point, really, in having … [PAUSES, as someone passes the table carrying a pint of beer] *… What's the point in having a job like mine, right, if you can't take the odd bribe, OK?* [LAUGHS] *… I mean … but anyway, … the most important thing is I'm pretty sure they can't prove it. I mean, it was all in cash, right, and if the police had any real evidence they'd have used it by now, right? … And they certainly won't find the money, that is safely tucked away in a safety deposit box in Jersey …* [LAUGHS] *until… well, until this all blows over."*

Experiment 3

This two-minute video features Amelia Palmer, a fictional 27-year-old social media influencer, sitting in a picture-perfect cottage kitchen. She is wearing a simple dress, silver pendant and make-up. Behind her on the kitchen counter is a light box displaying the message "LOVE VEGAN LIFE" to her right, with fresh fruit and vegetables carefully arranged on a vintage wooden chopping board to her left. She looks straight into the camera with a serious expression.

"Hello everyone. I wanted to come here today and talk to you about something that I think I need to address. Umm… obviously this is quite serious and I've upset quite a few people. And so I think it is time that I put the rumours and the whispers to bed, and tell you really what happened. … I am a vegan, and I've always been a vegan, and I've always stood for what I believe in. However, a few weeks ago, at a press event, somebody saw me in a restaurant, eating a burger. … I know that this might sound silly to some people, but to my followers, and to me, this is a really big deal. And I've let you all down. … I wanted to tell you a little bit more about where I was, where my headspace was at, in that moment. … I was in a bad place and I'd come to a point in my life where I was confused about what I was really truly standing for. … But seeing how much I've let you down and disappointed you, how many people have got in touch to say that, the way that they see me and what I do now is, just, false and wrong and a lie, it's really hurt. … And I can only imagine how much that's hurt you. … So all I can say, really, is I'm sorry. I apologise from deep within me, and I want you to know that I've learnt. And if anything, this experience, that day, the way that you have come to me to tell me how it's made you feel, that's made me even more passionate about being a vegan. … We should stand together for what we believe in. … Thank you."

Extended Results

Unreported Variables

Two additional variables measured in Experiment 1 are reported here.

Perception of suitability was measured with the question *"Do you think John Carter is a suitable person to hold a public position (e.g., planning department, tax office)?"*, using a 7-point Likert scale (–3 = *definitely not*, +3 = *definitely yes*), reverse coded to show unsuitability. Mean perception of unsuitability was 0.13 (*impossible to say*) for the *control* condition, and 1.45 (*maybe yes*) for the *fake specific* condition, $t(152) = 6.61$, $p < .001$, $d = 1.02$.

Perception of authenticity was measured with the question *"Do you think the video shows what actually happened?"*, using the same Likert scale, reverse coded to show inauthenticity. Mean perception of inauthenticity did not differ between conditions, $t(156) = 0.68$, $p = .248$, $d = 0.11$. This question was intended to test whether participants believed the video had been manipulated, but qualitative responses indicated that the question was widely misunderstood. Many participants referred to the sound effects which had been added to obscure incriminating content in the *control* condition (e.g., *definitely yes* / *"Most of it was bleeped out"*). Some participants stated that a description of something happening is not the event itself (e.g., *probably yes* / *"It may have told what happened but not shown"*). Other participants answered in relation to the accuracy of what was said (e.g., *maybe not* / *"While he said he accepted money I am not clear exactly what actually happened"*). We did not consider the results of this variable to be meaningful, given the various ways in which the question was interpreted by participants.

Preregistered Hypotheses

Experiment 1

H1 was supported: perception of guilt was greater in the *fake specific* condition (M = 0.87) than in the *control* condition (M = –0.40), $t(134) = 6.16$, $p < .001$, $d = 0.96$.

H2 was supported: perception of falsity was greater in the *fake specific* condition (M = 1.27) than in the *control* condition (M = –0.07), $t(156) = 6.07$, $p < .001$, $d = 0.93$.

Experiment 2

H1 was supported: perception of guilt was greater in the *fake specific* condition (M = 0.62) than in the *control* condition (M = –0.29), $t(55) = 2.60$, $p = .012$, $d = 0.60$.

H1a was unsupported: perception of guilt did not differ between the *fake generic* condition (M = 0.73) and the *fake specific* condition (M = 0.62, $p > .999$).

H1b was supported: perception of guilt was greater in the *fake none* condition (M = 2.14) than in the *fake generic* condition (M = 0.73), $t(61) = 4.22$, $p < .001$, $d = 0.95$.

H2 was supported: perception of falsity was greater in the *fake specific* condition (M = 1.88) than in the *control* condition (M = 0.00), $t(76) = 5.44$, $p < .001$, $d = 1.21$.

H2a was supported: perception of falsity was greater in the *fake specific* condition (M = 1.88) than in the *fake generic* condition (M = 0.18), $t(72), = 4.81$, $p < .001$, $d = 1.12$.

H2b was unsupported: perception of falsity did not differ between the *fake generic* condition (M = 0.18) and the *fake none* condition (M = 0.27, $p > .999$).

H3a was supported: perception of guilt was greater in the *real none* condition (M = 2.18) than in the *real generic* condition (M = 1.49, *maybe yes*), $t(63) = 2.62$, $p = .011$, d = 0.60.

H3b was unsupported: perception of guilt did not differ between the *real generic* condition (M = 1.49) and the *real specific* condition (M = 1.29, p > .999).

H4a was unsupported: perception of falsity did not differ between the *real generic* condition (M = –0.18) and the *real none* condition (M = 0.33, p > .999).

H4b was supported: perception of falsity was greater in the *real specific* condition (M = 0.97) than in the *real generic* condition (M = –0.18), $t(73) = 3.64$, $p < .001$, $d = 0.83$.

Experiment 3

H1 was supported: perception of guilt was greater than zero in the *fake specific* condition (M = 1.04), but lower than in the *real none* condition (M = 2.00), $t(89) = 2.74$, $p = .007$, $d = 0.53$.

H2 was supported: mean perception of falsity was greater in the *fake specific* condition (M = 0.76) than in the *real none* condition (M = –0.60), $t(76) = 5.44$, $p < .001$, $d = 1.21$.

H3 was unsupported: perception of guilt did not differ between the *real none* condition (M = 2.00) and the *real specific* condition (M = 1.59, $p = .119$).

H4 was unsupported: perception of falsity did not differ between the *real specific* condition (M = –0.39) and the *real none* condition (M = –0.60, $p = .518$).