Changing Beliefs about Past Public Events with Believable and Unbelievable Doctored Photographs

Robert A. Nash

School of Life and Health Sciences, Aston University

Word count: 6684 words (main text & footnotes)

Author note. The author is grateful to Harriet Adams, Helen Pericleous, and Anna Thompson for their assistance with data collection and the preparation of materials.

* Corresponding author

Robert A. Nash
School of Life and Health Sciences
Aston University
Birmingham B4 7ET
United Kingdom
Tel: +44 (0) 121 204 4522
Email: R.Nash1@aston.ac.uk

Abstract

Doctored photographs can shape what people believe and remember about prominent public events, perhaps due to their apparent credibility. In three studies, subjects completed surveys about the 2012 London Olympic torch relay (Experiment 1) or the 2011 Royal Wedding of Prince William and Kate Middleton (Experiments 2-3). Some were shown a genuine photo of the event; others saw a doctored photo that depicted protesters and unrest. A third group of subjects saw a doctored photo whose inauthenticity had been made explicit, either by adding a written disclaimer (Experiment 1) or by making the digital manipulation deliberately poor (Experiments 2-3). In all three studies, doctored photos had small effects on a subset of subjects' beliefs about the events. Of central interest though, comparable effects also emerged when the photos were overtly inauthentic. These findings suggest that cognitive mechanisms other than credibility—such as familiarity misattribution and mental imagery—can rapidly influence beliefs about past events even when the low credibility of a source is overt.

Key words: belief distortion; photographs; credibility; familiarity; images

Changing Beliefs about Past Public Events with Believable and Unbelievable Doctored Photographs

In 2011, a blatantly fake photograph of three government officials—whose feet seemed not to be touching the ground—appeared on a Chinese government website. The photo went viral, leading two New York Times bloggers to report "China admits officials cannot levitate" (Mackey & Harris, 2011). Their blog post quoted one Chinese citizen, who complained "Even a rank amateur like myself can tell that this was a Photoshop job, and they had the nerve to put this on the home page!". Doctored photos of diverse quality are common in the modern mass media, and there is a current resurgence of interest in people's ability to detect these forgeries as an instance of so-called 'fake news' (Edkins, 2016). Recent psychological studies show that people's ability in this respect is very often suboptimal (Nightingale, Wade, & Watson, in press); in fact, doctored images can sometimes lead people to believe in—and even remember—events that never occurred (e.g., Frenda, Knowles, Saletan, & Loftus, 2013; Sacchi, Agnoli, & Loftus, 2007). But does this rapid distortive influence only occur when viewers mistakenly treat the forgeries as genuine? The research reported here tackles this question.

Doctored photos as credible evidence

For decades, research has demonstrated numerous ways by which people's beliefs about their personal past can be altered (Loftus, 2005; Mazzoni, Loftus, & Kirsch, 2001). Many of these studies have relied on apparently highly credible evidence as a vehicle to deliver suggestive information. For example, studies into false autobiographical memories have recruited family members as credible sources of suggestion (Loftus & Pickrell, 1995; Ost, Foster, Costall, & Bull, 2005; Scoboria, Wysman, & Otgaar, 2012), whereas in typical misinformation studies, the credible sources are usually the experimenters themselves (Loftus & Palmer, 1974; Takarangi, Parker, & Garry, 2006). Another credible source of false information used in several studies is doctored photos (Garry & Gerrie, 2005). Such images are rather unique forms of suggestion, as they can provide seemingly authoritative 'proof' of fictional events' occurrence. Consequently, seeing credible digital forgeries can lead people to report altered attitudes toward branded products, to falsely remember experiences that never occurred, to falsely internalize guilt for a prohibited act, and to snitch on innocent peers (Hellenthal, Howe, & Knott, 2016; Hessen-Kayfitz & Scoboria, 2012; Nash & Wade, 2009; Nash, Wade, & Lindsay, 2009; Wade, Garry, Nash, & Harper, 2010; Wade, Garry, Read, & Lindsay, 2002; Wade, Green, & Nash, 2010; Wright, Nash, & Wade, 2015).

As well as altering people's personal histories, doctored images can also influence people's beliefs about significant public events. In one study, researchers doctored a photo of a peaceful protest in Rome, adding photographic details to falsely suggest that unrest and violence had occurred (Sacchi et al., 2007). Exposure to this doctored photo changed the way Italian subjects remembered the events that had occurred. Compared to subjects who saw the genuine photos, those who saw the doctored version reported that the protest had been more confrontational, violent, and negative, and that more damage, injuries and even deaths had occurred. Similarly, misled subjects said they would be less eager to participate in similar demonstrations in future. In a follow-up, Frenda et al.'s (2013) subjects saw doctored photos of fabricated political events, such as President Obama shaking hands with the former Iranian president, or President George W. Bush on vacation with a famous baseball player in the aftermath of Hurricane Katrina. After viewing the photos, nearly half of people said they remembered the false events happening—especially those events that suited their own political agendas.

How important is credibility?

Sacchi et al. (2007) proposed that the apparent authenticity of doctored photos is a crucial ingredient in their potency, perhaps because this authenticity makes people less likely to immediately reject them as false, and instead to more readily elaborate on what they might have seen. In line with this account, there is abundant evidence that people's susceptibility to suggestion depends on what or whom they believe the source of the false information to be (Dodd & Bradshaw, 1980; French, Garry, & Mori, 2011; Scoboria et al., 2012). These studies indicate that for suggestions to influence our cognitions about the past, they must normally be understood to originate from credible, trustworthy sources (see Nash, Wheeler, & Hope, 2015 for further discussion).

Yet despite strong evidence for the role of credibility, numerous other studies show us that information does not always need to be authoritative to shape what we believe. This point is demonstrated notably in studies of the illusory truth effect, where repeated exposure to purported 'facts' leads people, over time, to increasingly accept those facts as truth, even when the source of the facts is explicitly identified as unreliable (Henkel & Mattson, 2011), and when the facts are initially known to be untrue (Fazio, Brashier, Payne, & Marsh, 2015). Similar acceptance of misinformation in lieu of source credibility has also been observed in studies that involve photographic sources. Indeed, in Nash, Wade, and Brewer's (2009) study, a credibility mechanism alone accounted for most of the effect of doctored video-recordings on people's beliefs, but could not explain the entire effect. Those authors concluded that part of the effect of photographic images lies in their ability to evoke feelings of familiarity with the events they depict. In line with this conclusion, Newman and colleagues have found that when people evaluate trivia claims, they are more likely to judge those claims as true if the claims are shown alongside topic-relevant photos that provide no probative evidence (Newman, Garry, Bernstein, Kantner, & Lindsay, 2012; Newman et al., 2015). In earlier studies, similarly nonprobative photos have induced false memories of news headlines (Strange, Garry, Bernstein, & Lindsay, 2011), recent personal experiences (Henkel, 2011), and childhood experiences (Lindsay, Hagen, Read, Wade, & Garry, 2004).

Photos can be influential, then, even when they are not treated as credible evidence of what happened. Yet this may be true only when those photos seem incidental and innocuous, rather than when they can be explicitly identified as a source of misinformation. In these latter cases, the low credibility of the photo may alert viewers to avoid heuristic cues to truth. For example, we know from several other studies that false beliefs are less likely to occur when people are vigilant to suggestive influence, such as when they are forewarned that they are about to see information that may mislead them (Gallo, Roediger, & McDermott, 2001; Gerrie & Garry, 2011; Greene, Flynn, & Loftus, 1982). Based on these findings, we might expect that when people see photos that are overtly fake, they would be vigilant to possible influence and would therefore resist being misled by those photos. At present, we do not know whether this is the case, or whether other mechanisms can still lend influence to doctored photos even when their credibility is low. Indeed, it is noteworthy that even when forewarned about misleading information, people's susceptibility to suggestive influence is not typically eliminated even if it is reduced (see also Chambers & Zaragoza, 2001; Harris, 1978; Harris, Teske, & Ginns, 1975). Based on the literature reviewed, it was therefore hypothesized that undermining the credibility of a doctored photo would similarly reduce, but not entirely eliminate, its effects on people's beliefs. In the present research, Sacchi et al.'s (2007) doctored-photo methodology was used to test this hypothesis.

Experiment 1

One straightforward way to undermine the credibility of doctored images is to overtly identify them as fakes. For instance, women in one study who were exposed to digitally-idealized photos of fashion models reported greater body satisfaction when those photos were explicitly labelled as altered (Slater, Tiggemann, Firth, & Hawkins, 2012). The purpose of Experiment 1 was to search for analogous effects in the domain of beliefs about public events, by adding a disclaimer to the doctored photo seen by some subjects.

Method

Subjects and Design. A total of 203 UK-resident adults (76.8% females) aged 18-61 (*M*= 24.8, *SD*= 8.8) completed the study either online (*n*= 140) or on paper (*n*= 63) without compensation. In this and the subsequent studies, sample sizes were based solely on the number of people who took part within predetermined testing periods. Each subject was randomly assigned to either the Genuine (*n*= 67), Doctored (*n*= 65) or Doctored-disclaimer (*n*= 71) photo condition, and informed that the study was about 'memory for public events'.[1]

Materials and Procedure.

Photos. A recent and prominent target event was chosen for this study, namely the torch relay that preceded the London 2012 Olympic Games (the data were collected during 2013). The chosen target photo, obtained from the Internet, was in color and depicted a relay-runner carrying the Olympic torch past cheering crowds (the *Genuine* photo). For the *Doctored* condition, a professional graphic artist manipulated this photo by adding anti-Olympic placards amongst the crowd, adding numerous riot police officers, and changing the runner's face to appear worried rather than jubilant. For the *Doctored-disclaimer* condition, a prominent text-box was added to the upper-left corner of the doctored photo, stating 'Note: This image has been Photoshopped'.[2]

Questionnaire. Subjects completed the questionnaire at their own pace. After consenting to participate and providing demographic details, subjects were shown one of the three photos and were asked to write in their own words which event they thought the photo portrayed. The purpose of this question was merely to ensure that subjects paid attention to

[1] The questions and the visual layout of the 'pages' were identical across presentation formats (online vs. paper). Mann-Whitney tests revealed no significant differences between presentation formats for any target question with the exception of the Arrests question, where online subjects made significantly higher estimates than those who completed paper-based surveys. The proportions of subjects who completed the study in paper format was closely matched across experimental conditions (Genuine = 33%; Doctored = 29%; Doctored-Disclaimer = 31%; χ^2= 0.20, *p*= .91), therefore in all analyses, the data are collapsed across presentation formats.

[2] For copyright compliance reasons, the photographs used in these experiments cannot be reproduced here; however they can be obtained from the author on request.

the content of the photograph, regardless of whether or not they identified the event correctly (95.6% mentioned the Olympics and/or the torch or torch relay). On the subsequent page, the same image was shown again beneath the caption 'The event depicted in this photo is the 2012 Olympic torch relay.' Subjects rated how well they remembered the event (1= not at all, 5= very well).

On the next page, the photo was removed and subjects were instructed to answer the remaining questions based solely on their memory of the event. After one filler question regarding the number of people who watched the torch relay, they answered the target questions.

[TABLE 1 ABOUT HERE]

Target questions. All the verbatim target questions and scale anchors are listed in Table 1. Like Sacchi et al. (2007), we asked questions about details of the event itself, plus additional questions about subjects' attitudes and behavioural intentions. The choice of target questions was thematically informed by, but not identical to, the questions used by Sacchi et al. Subjects were first asked questions about the event itself, beginning by estimating the proportion of spectators at the event who were protesters rather than supporters. Like in Sacchi et al. (2007), this was the critical target question for which significant effects of doctored photos were predicted. All additional questions were selected based on exploratory rather than *a priori* theoretical reasons, and continued with estimates of the number of arrests made, and the number of injuries (note that among those who completed the survey in paper format, up to three data points were missing per condition for these three variables; see Table S1 in supplemental materials). Subjects also rated on 5-point scales how violent the protesters were, how much damage was caused to property, how successful the police were in controlling the crowds, and whether the relay went as smoothly as expected. Subjects were next asked questions about their future intentions: they rated on 5-point scales how likely they would be to attend a similar event in the future if it occurred, and the likelihood they would protest a similar event if it occurred. Before being debriefed, subjects were invited to write any comments they had about the study.

Results

Overall, subjects reported remembering the Olympic torch relay quite well ($M = 3.73$ out of 5, $SD = 1.23$). Looking to the main research question, the data differed substantially from normal distributions for most target variables, and transforming the data did not resolve this issue. Nonparametric Kruskal-Wallis tests are therefore reported for each of the five dependent variables that met the assumption of homogeneity of variances, as assessed via a nonparametric Levene's test (Nordstokke & Zumbo, 2010). For the remaining four dependent variables that did not meet this assumption, Mood's median χ^2 test was instead calculated. As the middle columns of Table 2 show, these analyses revealed significant differences between conditions for three of the nine target questions. In particular on the critical question, subjects' estimates of the proportion of spectators who were protesters differed significantly across conditions, as did their ratings of how smoothly the event went, and of the police's success in controlling the crowds. Follow-up pairwise Dunn tests or Mood's median χ^2 test (as appropriate) for these three variables show that responses in the Genuine condition differed significantly from those in the Doctored condition (all ps< .05; see rightmost columns of Table 2), but also from those in the Doctored-disclaimer condition (all ps< .05). Responses did not differ between the Doctored and Doctored-disclaimer conditions for any of these variables (all ps> .10). Responses did not differ significantly across conditions for any of the other target questions.[3]

[TABLE 2 ABOUT HERE]

[3] In this and the subsequent experiments, follow-up analyses were also conducted that excluded those subjects who rated their recollection of the target events as '1' (not at all). During revision of this paper, these analyses were removed in favor of the mini-metaanalytic approach outlined below.

Experiment 2

The results of Experiment 1 suggest that doctored photos influenced some of the beliefs held by subjects about the torch relay, and that a disclaimer was insufficient to negate this effect. This finding might imply that the credibility of a doctored photo is not a prerequisite for its capacity to distort beliefs. However, even though the photos in the Doctored-disclaimer condition were explicitly identified as doctored, it is plausible that some subjects in that condition still treated them as authentic. For example, they may not have noticed the disclaimer. In a small pilot study, 32 volunteers viewed the same materials as the online Doctored-disclaimer subjects from Experiment 1, but rather than receiving the target questions after viewing the photo for the second time, they were instead asked to report what the disclaimer in the corner of the photo has said. In total, 27 of the volunteers (84%) were able to repeat the gist, which supports the idea that most subjects in Experiment 1, but perhaps not all, would have noticed the disclaimer.

Yet even if they did notice the disclaimer, subjects may still have not understood the word ‘Photoshopped’, or may have presumed that details irrelevant to the authenticity of the image had been manipulated, such as the coloring. To counter these interpretations and to develop these findings further, in Experiment 2 the credibility of doctored photos was undermined in a different way – by making the quality of the manipulation deliberately obvious. This variation meant that subjects would be highly likely to know that aspects of the doctored photo were fake. They might not notice all of the manipulations, of course, but should nevertheless be clear that the image is not entirely authentic. This time, subjects also saw photos of a more discrete target event (i.e., an event that occurred on one single day), namely the 2011 Royal Wedding of Prince William and Kate Middleton.

Method

Subjects and Design. A total of 174 UK-resident adults (69.0% females) aged 18-72 (*M*= 27.8, *SD*= 11.8) participated online during 2013 without compensation. Each subject was randomly assigned to either the Genuine (*n*= 57), Well-doctored (*n*= 58) or Badly-doctored (*n*= 59) photo condition, and they were told that the study was about memory for public events.

Materials and Procedure.

Photos. A color photo was obtained from the Internet, depicting the Royal couple driving alongside cheering crowds on their wedding day (the *Genuine* photo). A graphic artist created two doctored versions of this photo, adding anti-monarchy protesters holding placards, as well as extra police officers on foot, horseback, and motorcycle. For the *Well-doctored* condition, great effort was invested in making this manipulation convincing. For the *Badly-doctored* condition, the manipulation involved the same additions but was deliberately unconvincing – objects were the wrong size, wrong color, and/or wrong proportions, some were coarsely pixelated, the police motorcycle was missing its front wheel, and so forth.

As a manipulation check, 60 volunteers were shown five doctored filler photos collected from the Internet (including, of course, one of levitating Chinese government officials), followed randomly by a target photo: either the Genuine, Well-doctored, or Badly-doctored photo of the Royal Wedding. Subjects were told that all six photos had been doctored, and they rated the convincingness of each (1= not at all convincing; 7= very convincing). Their ratings of the target photo differed by condition, $F(2, 57)= 17.25$, $p< .001$; post-hoc comparisons showed that the Well-doctored photo ($M= 4.55$, $SD= 1.99$) was indeed more convincing than the Badly-doctored photo ($M= 2.55$, $SD= 1.67$, $p< .01$), although not as convincing as the Genuine photo ($M= 5.90$, $SD= 1.77$, $p= .03$).

Questionnaire. Similarly to Experiment 1, subjects in the main part of Experiment 2 were first shown one of the three photos and asked which event they thought the photo portrayed (93.7% mentioned the Royal Wedding, though most of the remaining 6.3% also gave responses that indicated clear engagement with the photo, e.g., "Anti monarchy protest" or naming the wrong members of the royal family). They were not told anything about the authenticity of the photo. On the subsequent page, the same image was re-shown beneath the caption 'The event depicted in this photo is the 2011 Royal Wedding of Prince William and Kate Middleton'. The remaining procedure mirrored Experiment 1 except for some different target questions; no filler questions were included.

Target questions. The verbatim target questions and scale anchors are listed in Table 1, and were again thematically informed by, but not identical to, the questions used by Sacchi et al. (2007). Like Experiment 1, these began with questions about the event itself. Subjects first answered the critical question, estimating the proportion of spectators at the event who were protesters rather than supporters. They then estimated the number of arrests made, and the number of extra police on duty, and they rated on a 5-point scale how violent the protesters were. Next they were asked questions about attitudes – they rated on 5-point scales how supportive both they and the British public were of the wedding, and how positive both they and the British public thought the event was. Subjects then answered questions about their intentions: they rated on 5-point scales how likely they would be to attend a similar event in the future if it occurred, and the likelihood that they would protest a similar event if it occurred. Before being debriefed, subjects were invited to add comments about the study.

Results

Overall, subjects reported remembering the Royal Wedding quite well (M = 3.24 out of 5, SD = 1.20). Due to non-normal data distributions, Kruskal-Wallis tests were again conducted for each dependent variable, with the exception of the 'Protest in future' variable, which did not meet the assumption of homogeneity of variances, and for which Mood's median χ^2 test was instead conducted. As the middle part of Table 3 shows, there were significant differences across conditions for three of the ten target questions. As in Experiment 1, subjects' estimates of the proportion of spectators who were protesters differed substantially across conditions. So too did their estimates of the number of arrests made, and their ratings of how supportive the British public were of the Royal Wedding. Follow-up pairwise Dunn tests for these three variables showed that responses in the Genuine condition differed significantly from those in the Well-doctored condition (ps< .04; see rightmost columns of Table 3), with the exception of the number of arrests, which was not significant, p= .10. The responses in the Genuine condition also differed significantly from those in the Badly-doctored condition for all three of these variables (all ps< .01), but the Well-doctored and Badly-doctored conditions did not differ significantly for any of the three (ps> .34). Responses did not differ significantly across conditions for any of the other target questions.

[TABLE 3 ABOUT HERE]

Experiment 3

The findings of Experiment 2 give us greater confidence that subjects were influenced even by photos that they knew were not wholly authentic. Experiment 3 set out to replicate these findings, but with some additional measures inspired by Laney et al.'s (2008) *Red Herring Technique*, to help rule out the possibility that the results could have been driven by demand effects – important because neither Experiment 1 nor 2 included systematic checks on subjects' suspicion. In the Red Herring Technique, subjects are given a false cover story about the experiment's purpose just like in Experiments 1 and 2. But in addition, several 'foil' questions are inserted into the study procedure, designed to lead subjects to suspect that the experiment is actually about some other, equally false, topic. These red herrings reduce the chance that subjects who are suspicious about the experiment's intent will alight on the actual hypothesis. In Experiment 3 then, subjects were informed that the study was about 'attitudes towards monarchy and patriotism,' but throughout the questionnaire they answered several questions that could lead them to think the study was really about public confidence in policing and security. To measure the effectiveness of this ploy, at the end of the study subjects were asked what they believed the aim of the experiment to be.

Method

Subjects and design. A total of 217 UK-resident adults (46.1% females) aged 18-84 (M= 31.3, SD= 16.6) participated online during 2014 without compensation. Each was randomly assigned to either the Genuine (n= 74), Well-doctored (n= 79) or Badly-doctored (n= 64) photo condition.

Materials and Procedure.

Photos. The photos used in this study were identical to those used in Experiment 2, except for the badly-doctored photo. It was clear from the manipulation check in Experiment 2 that people already found the badly-doctored unbelievable; nevertheless it seemed there was some room to make this photo even less believable and thus to strengthen the credibility manipulation further. Efforts were therefore taken to this end; specifically, several of the false details that had been digitally added to this photo were given colored border-edging that mismatched their surrounding. This manipulation gave the distinct impression that these details had been poorly ‘cut out’ from another image. Using the same manipulation check as in Experiment 2, 60 new volunteers rated the convincingness of six doctored photos, including one of the three target photos. Their ratings of the target photo differed by condition, $F(2, 57)$= 69.14, p< .001; post-hoc comparisons showed that the Well-doctored photo (M= 5.90, SD= 1.25) was far more convincing than the Badly-doctored photo (M= 2.10, SD= 1.25, p< .001), and almost as convincing as the Genuine photo (M= 6.25, SD= 1.21, p= .38).

Questionnaire. Subjects completed the same questionnaire as in Experiment 2, except for four changes. The first was that seven filler questions were intermixed amongst the target questions, two of which related to the stated cover story of the study (i.e., about patriotism) and five formed the ‘red herring’ (about policing). For example, subjects were asked “To what extent would you say that you are ‘proud to be British’?” and “How much authority in general do you believe the police have in the UK?”. The second minor change was that the target question about how many extra police were on call during the Royal Wedding was removed, because the variability of the data from Experiment 2 indicated that people struggled to estimate this number, and it was replaced with a question about police control, like in Experiment 1. Third, all questions using Likert scale responses used 7-point, rather than 5-point scales, to permit greater sensitivity in subjects’ responses. The fourth change was that before being debriefed, subjects were required to write in their own words what they believed the aim of the study to be.

Results

Guessing the study’s aims. Before analyzing the main data, subjects’ ideas about the study’s aims were first inspected. A research assistant coded each of these, blind to condition, as either (1) consistent with the stated cover story about monarchy and patriotism; (2) consistent with the red herring, mentioning details about policing and/or security (either alone, or alongside references to monarchy and patriotism); (3) consistent with the true aim of the study, mentioning the suggestiveness of (doctored) images; or (4) some other unrelated idea.

The majority of subjects (53.4%) were apparently unsuspicious, suggesting study aims that were consistent with the cover story they received. A further 36.4% were coded as ‘red herring’ responses, which suggests that this ploy was successful. Indeed, some of these subjects’ responses confirm that they had been eager to identify ulterior aims of the study; for example one wrote “[A psychology student] once told me they are outright liars in the titles of their studies otherwise it affects their outcome. I'd therefore guess that the study is about public opinions towards police and their methods.” Only four subjects appeared to have a sense of the true purpose of the study, although a fifth subject who might have been suspicious (“Use of pictures/media to influence public patriotism”) was added to this category, to be conservative (2.3%). The remaining 7.8% of subjects made unrelated guesses or indicated that they had no idea.

In short, these data indicate that awareness of the study’s aims likely played a very minimal role in the present findings. The five subjects who were coded as correctly guessing the study’s aims were removed from analyses; these analyses are therefore based on the remaining 212 subjects, of whom 94.3% mentioned the royal wedding when asked which event the photo portrayed (like in Experiment 2, most of the remainder also gave responses that indicated clear engagement with the photo). These subjects said that they recalled the wedding quite well (M = 4.70 out of 7, SD = 1.89).

Main analysis. Kruskal-Wallis tests were again conducted for each dependent variable except for the ‘Protest in future’ variable, which did not meet the assumption of homogeneity of variances, and for which Mood’s median χ^2 test was instead conducted. As the middle part of Table 4 shows, there were significant differences for two of the nine target questions, including the critical question. Subjects’ estimates of the proportion of spectators who were protesters differed across conditions, as in both previous experiments, and their estimates of the number of arrests made also differed. Follow-up pairwise Dunn tests showed that for both of these variables, estimates in the Genuine condition were greater than those in the Well-doctored condition (both ps< .05; see rightmost columns of Table 4) and the Badly-doctored condition (ps< .03). Responses did not differ between the Well-doctored and Badly-doctored conditions for either variable (ps> .17). No other statistically significant differences emerged for any other target question.

[TABLE 4 ABOUT HERE]

Effect size estimation

Of the 14 different target questions used across Experiments 1-3, ten were measured in more than one experiment, and five (including the critical question) were measured in all three experiments. It is therefore useful to statistically combine data across experiments, to gain more precise estimates of the effects of the believable and unbelievable fake photos on people's beliefs, rather than relying solely on null hypothesis significance testing (Cumming, 2012). The so-called 'mini meta-analysis' is an excellent tool for this task (Goh, Hall, & Rosenthal, 2016).

Method

Using the procedure outlined by Ruscio (2008), the effect-size measure *A* was calculated for each target variable measured in each of Experiments 1 to 3, for each of the three pairwise contrasts of experimental conditions. This effect size measure, as Ruscio has demonstrated, is highly robust to violations of parametric assumptions, and can be easily converted to a standardized measure d_A using the procedure described by Li (2016). These d_A measures were therefore calculated in order to permit the straightforward combination of effect sizes across experiments. Estimates of d_A are reported in full in Table S1 in the online supplementary materials, and can be interpreted as standardized mean differences equivalent to Cohen's *d*.

For the ten variables that were measured in multiple experiments, d_A values were combined across experiments via a series of random effects mini meta-analyses. For the purposes of these analyses, the 'Doctored' condition of Experiment 1 and the 'Well-doctored' conditions of Experiments 2 and 3 were treated as equivalent, and represent the *Believable fake* conditions. Likewise, the 'Doctored-disclaimer' condition of Experiment 1 and the 'Badly-doctored' conditions of Experiments 2 and 3 were treated as equivalent, and represent the *Unbelievable fake* conditions. These analyses were conducted both using 95% confidence intervals and, to account for the inflated likelihood of Type I errors, using more conservative

99.9% confidence intervals (calculated using ESCI; Cumming, 2012). For the four variables measured in only one experiment, confidence intervals for d_A were again calculated, but those results are not discussed further.

Results

The results of these analyses are reported in Table 5, and three specific findings warrant particular attention. First, and most importantly, combining data across the three experiments revealed that subjects' estimates of the prevalence of protesters were significantly inflated both by believable and unbelievable fake photos. Indeed, both kinds of photo had medium-sized effects on these prevalence estimates, relative to genuine photos. Although unbelievable fakes appeared to have a somewhat smaller influence on these estimates than did believable fakes, the 95% confidence interval for this difference marginally included zero. All of these conclusions held even after reducing the likelihood of Type I error by using 99.9% confidence intervals.

[TABLE 5 ABOUT HERE]

Second, across the three experiments, subjects' estimates of the number of arrests were also significantly boosted both by believable and unbelievable fakes (relative to genuine photos), and these effects were in the order of small-to-medium. As Table 5 shows, the significant effect of unbelievable fakes held even after reducing the likelihood of Type I error, but in this more conservative analysis the 99.9% confidence interval for the effect of believable fakes marginally included zero.

Third, across the three experiments, believable fakes led subjects to estimate greater levels of violence at the public events; this meta-analytic effect was in the order of small-to-medium. However, this effect did not remain statistically significant after reducing the likelihood of Type I error. In both analyses, the effect of unbelievable photos did not differ significantly either from the effect of genuine photos, or of believable photos.

As Table 5 shows, there were no significant differences between conditions for any of the other seven variables that were measured in multiple experiments. In sum, when the data were combined across experiments, there was clear evidence that both believable and unbelievable fake photos influenced subjects' beliefs about the critical question, namely the prevalence of protesters at the target events. There was also some evidence that other elements of people's beliefs about these events were influenced similarly, but no strong evidence that the effects of doctored photos were reliably reduced by undermining their credibility.

General Discussion

These data represent the latest empirical demonstration of the power of fabricated images to alter people's beliefs about public events (Frenda et al., 2013; Sacchi et al., 2007). In three experiments, doctored photographs shaped certain beliefs about two relatively recent events of particular salience to the British public – the 2012 Olympic Torch relay, and the 2011 Royal Wedding. Most prominently, there were consistent effects on the critical target question: in all three studies, the digital forgeries led subjects to believe that the prevalence of protesters among the spectators at these events was greater. What is more striking is that the initial hypothesis—that undermining source credibility would reduce but not eliminate the effects on doctored photos—was only partially supported. The effects were certainly not eliminated, but in fact, they were not reliably reduced either. In other words, this medium-sized effect occurred even when the photo's inauthenticity was overt, either due to an explicit written disclaimer, or because the digital manipulation was poor. Such findings support the conclusion that mechanisms aside from credibility can lend influence to doctored images even when their credibility is low. This point will be discussed shortly.

Alongside the critical target question, the believable and unbelievable doctored photos also seemed to have somewhat weaker effects on responses to other target questions. Notably, there was some evidence that both kind of doctored photos led subjects to estimate that significantly more people were arrested at the target events. Believable fakes also increased subjects' estimates of violence, although it is difficult to interpret whether undermining the credibility of these photos did diminish this effect, or indeed whether it could be attributed to Type I error due to the large number of variables measured. Likewise, other significant effects of believable and unbelievable doctored photos emerged within the individual studies, but the replicability and robustness of those individual findings is unclear.

In general, statistically significant effects were rather rarer than one might expect based on Sacchi et al.'s (2007) research, in which statistically significant effects of doctored photos emerged from 16 of 27 inferential tests of similar target questions across two experiments. Nevertheless, the analyses here support the conclusion that at least some effects of believable and unbelievable doctored photos were robust, and not easily attributed to the inflated likelihood of Type I errors. The fact that the effects were strongest for the critical question—estimates of the prevalence of protesters—is likely related to the fact that the doctored photos themselves depicted protesters; these false beliefs were therefore at least partly consistent with the literal content of the photos, whereas any effects on other target questions (e.g., to believe that there was violence) would require a larger leap of inference.

The influence of unbelievable doctored photos might seem counterintuitive, but it is worth noting that even in Frenda et al.'s (2013) research, the doctored photos did not always constitute credible proof of the suggestions. For instance, their photo might seem to prove that George W. Bush spent time with the baseball star, yet it contained no evidence that this happened specifically during the aftermath of Hurricane Katrina. That particular element of the suggestion was provided in an accompanying written caption, which transformed an otherwise innocuous doctored photo into one that could be politically toxic. In short, that particular doctored photo might well have led people—especially Liberals—to believe they recalled the fictional Bush story, but this was not because the photo contained credible proof of the politically damaging details.

If credibility is not wholly responsible for the effects of doctored photos, then what other mechanisms might play a role? In some cases, people accept suggestions from unbelievable sources because, after a sufficient delay, they continue to remember the misinformation but forget the cues concerning low source credibility (Underwood & Pezdek, 1998). This 'sleeper effect' seems an unlikely explanation here, given that subjects in these studies—like Sacchi et al.'s (2007) and Frenda et al.'s (2013) subjects—reported their beliefs only moments after seeing the doctored photos. It also seems unlikely that demand effects can explain these findings. First, like in Sacchi et al. (2007), several subjects in Experiments 1 and 2 made spontaneous comments suggesting they were not aware of having been misled. For example, one wrote "Strange that people came to protest, being violent. Unsure of the study aims". Second, and more powerfully, in Experiment 3 subjects were formally prompted to report their ideas about the study's aims, and a red herring was used to disguise the true aims. In these circumstances, still very few subjects guessed correctly, and even after excluding those guessers some effects of both well-doctored and badly-doctored photos still obtained.

At least two other mechanisms are conceivable, although further studies would be necessary to test these mechanisms directly. First, these findings are consistent with the idea that photos provide a degree of familiarity to suggestions, even when they offer no probative evidence (Nash, Wade, & Brewer, 2009; Newman et al., 2012; Strange et al., 2011). According to this explanation, simply seeing images of protesters and police—irrespective of how convincingly they were embedded into the scene—could have led subjects to process thoughts about aggression and unrest more fluently, with this fluency being misattributed to familiarity and prior experience (Brown & Marsh, 2009; Fazio et al., 2015; Jacoby et al., 1989; Winkielman, Reber, Schwarz, & Fazendeiro, 2003). This explanation fits neatly with the finding that digitally adding police officers into photos of safe-looking environments can lead those environments to be judged as less safe (van de Veer, de Lange, van der Haar, & Karremans, 2012).

A second, but not mutually exclusive possibility is that the false details in the doctored photos prompted subjects to retrieve mental images of protests and violence at other events, in real-life or on television. Indeed, media speculation about the possibility of protests preceded both the Olympic torch relay and the Royal Wedding—although both events in reality occurred with minimal disruption (Gibson & Walker, 2012; Malik, 2011)—and subjects might have even imagined these hypothetical protests at the time of the media coverage. We know that photos can lead people to elaborate on the information they receive, and to then believe they actually saw or heard those elaborated details (e.g., Garry, Strange, Bernstein, & Kinzett, 2007; Henkel, 2012). Therefore, the images of unrest in the convincing and unconvincing doctored photos could have encouraged subjects to retrieve relevant mental imagery, which they might in turn have misattributed to their recollections of the target events via a feature importation mechanism (Henkel & Carbuto, 2008; Johnson, Hashtroudi, & Lindsay, 1993; Lyle & Johnson, 2006). Whether these, or other mechanisms are responsible for the effects we have observed, an important finding is that the effects obtained even though the low credibility of the source should have given subjects a strong steer to be vigilant to potential influence.

The findings of these studies have clear implications for the popular media, wherein digital fakes are published regularly and there is currently great interest in people's ability to detect and reject so-called 'fake news'. In particular, it is interesting to consider the influence that fake images might have when published in low-credibility media sources, and are therefore not expected to be taken seriously (see Kelly & Nace, 1994). In future studies it would be valuable to examine whether the distortive power of doctored photos grows as time elapses after exposure, and as any doubts about source credibility are gradually replaced with a hazy sense of familiarity (Brown & Marsh, 2009; Ecker, Lewandowsky, & Tang, 2010; Skurnik, Yoon, Schwarz, & Park, 2005). A further question not addressed here is to what extent people's general proclivity to trust photos—relative to written or spoken suggestions—plays a role in allowing even blatantly false suggestions to influence people's beliefs. For example, a habitual tendency to trust photos might mean people require more time to process source credibility cues, time during which rapid mechanisms such as familiarity and fluency could be influencing people's beliefs. Future research should examine the plausibility of this speculative account. Finally, given the increasing focus on the behavioral consequences of false beliefs, future studies might explore whether doctored images—believable ones, or blatant ones—ever lead people to change their behavior (Bernstein, Scoboria, & Arnold, 2015; Clifasefi, Bernstein, Mantonakis, & Loftus, 2013). An obvious forgery, it seems, could be influential just like a compelling forgery – something that might give pause for thought to those who mocked the photo of levitating Chinese officials.

References

Bernstein, D. M., Scoboria, A., & Arnold, R. (2015). The consequences of suggesting false childhood food events. *Acta Psychologica, 156*, 1-7.

Brown, A. S., & Marsh, E. J. (2009). Creating illusions of past encounter through brief exposure. *Psychological Science*, *20*, 534-538.

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition, 29*, 1120-1129.

Clifasefi, S. L., Bernstein, D. M., Mantonakis, A., & Loftus, E. F. (2013). "Queasy does it": False alcohol beliefs and memories may lead to diminished alcohol preferences. *Acta Psychologica*, *143*, 14-19.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning & Verbal Behavior, 19,* 695-704.

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*, 1087-1100.

Edkins, B. (2016, December 20). Americans believe they can detect fake news. Studies show they can't. *Forbes* Retrieved from www.forbes.com/sites/brettedkins/2016/12/20/americans-believe-they-can-detect-fake-news-studies-show-they-cant

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*, 993-1002.

French, L., Garry, M., & Mori, K. (2011). Relative – not absolute – judgments of credibility affect susceptibility to misinformation conveyed during discussion. *Acta Psychologica, 136*, 119-128.

Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology, 49*, 280-286.

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review, 8,* 579-586.

Garry, M., & Gerrie, M. P. (2005). When photographs create false memories. *Current Directions in Psychological Science, 14*, 321-325.

Garry, M., Strange, D., Bernstein, D. M., & Kinzett, T. (2007). Photographs can distort memory for the news. *Applied Cognitive Psychology, 21*, 995-1004.

Gerrie, M. P., & Garry, M. (2011). Warnings reduce false memories for missing aspects of events. *Experimental Psychology*, *58*, 207-216.

Gibson, O., & Walker, P. (2012, April 18). London 2012: Protesters urged to keep away from Olympic torch relay. *The Guardian.* Retrieved May 1, 2013 from www.guardian.co.uk/sport/2012/apr/18/london-2012-protesters-olympic-torch.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10*, 535-549.

Greene, E., Flynn, M., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning & Verbal Behavior, 21,* 207-219.

Harris, R. J. (1978). The effect of jury size and judge's instructions on memory for pragmatic implications from courtroom testimony. *Bulletin of the Psychonomic Society, 11*, 129-132.

Harris, R. J., Teske, R. R., & Ginns, M. J. (1975). Memory for pragmatic implications from courtroom testimony. *Bulletin of the Psychonomic Society, 6*, 494-496.

Hellenthal, M. V., Howe, M. L., & Knott, L. M. (2016). It must be my favourite brand: Using retroactive brand replacements in doctored photographs to influence brand preferences. *Applied Cognitive Psychology, 30*, 863-870.

Henkel, L. A. (2011). Photograph-induced memory errors: When photos make people claim they've done things they haven't. *Applied Cognitive Psychology, 25*, 78-86.

Henkel, L. A., (2012). Seeing photos makes us read between the lines: The influence of photos on memory for inferences. *Quarterly Journal of Experimental Psychology, 65*, 773-795.

Henkel, L. A., & Carbuto, M. (2008). How source misattributions arise from verbalization, mental imagery, and pictures. In M. Kelley (Ed.), *Applied memory* (pp. 213-234). Hauppauge, NY: Nova Science Publishers.

Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness & Cognition, 20*, 1705-1721.

Hessen-Kayfitz, J. K., & Scoboria, A. (2012). False memory is in the details: Photographic details differentially predict memory formation. *Applied Cognitive Psychology*, *26*, 333-341.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*, 326-338.

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

Kelly, J. E., & Nace, D. (1994). Digital imaging and believing photos. *Visual Communication Quarterly, 1*, 4-5 & 18.

Laney, C., Kaasa, S. O., Morris, E. K., Berkowitz, S. R., Bernstein, D. M., & Loftus, E. F. (2008). The Red Herring technique: A methodological response to the problem of demand characteristics. *Psychological Research*, *72*, 362-375.

Li, J. C. H. (2016). Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data. *Behavior Research Methods, 48*, 1560-1574.

Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science, 15*, 149-154.

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

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction. *Journal of Verbal Learning & Verbal Behavior, 13*, 585-589.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25,* 720-725.

Lyle, K., & Johnson, M. (2006). Importing perceived features into false memories. *Memory, 14*, 197-213.

Mackey, R., & Harris, E. A. (2011, June 30). China admits officials cannot levitate. Retrieved May 1, 2013 from http://thelede.blogs.nytimes.com/2011/06/30/china-admits-officials-cannot-levitate/.

Malik, S. (2011, May 1). Now the royal wedding activists say they were held by police to avert protests. *The Guardian.* Retrieved May 1, 2013 from www.guardian.co.uk/uk/2011/may/01/wedding-activists-rounded-up-police.

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.

Nash, R. A., & Wade, K. A. (2009). Innocent but proven guilty: Eliciting internalized false confessions using doctored-video evidence. *Applied Cognitive Psychology*, *23*, 624-637.

Nash, R. A., Wade, K. A., & Brewer, R. J. (2009). Why do doctored images distort memory? *Consciousness & Cognition, 18*, 773-780.

Nash, R. A., Wade, K. A., & Lindsay, D. S. (2009). Digitally manipulating memory: Effects of doctored videos and imagination in distorting beliefs and memories. *Memory & Cognition*, *37*, 414-424.

Nash, R. A., Wheeler, R. L., & Hope, L. (2015). On the persuadability of memory: Is changing people's memories no more than changing their minds? *British Journal of Psychology, 106*, 308-326.

Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review*, *19*, 969-974.

Newman, E. J., Garry, M., Unkelbach, C., Bernstein, D. M., Lindsay, D. S., & Nash, R. A. (2015). Truthiness and falsiness of trivia claims depend on judgmental contexts. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41*, 1337-1348.

Nightingale, S. J., Wade, K. A., & Watson, D. G. (in press). Can people identify original and manipulated photos of real-world scenes? *Cognitive Research: Principles and Implications.*

Nordstokke, D. W., & Zumbo, B. D. (2010). A new nonparametric Levene test for equal variances. *Psicologica, 31*, 401-430.

Ost, J., Foster, S., Costall, A. & Bull, R. (2005). False reports of childhood events in appropriate interviews. *Memory, 13,* 700-710.

Ruscio, J. (2008). A probability-based measure of effect size: robustness to base rates and other factors. *Psychological Methods, 13*, 19-30.

Sacchi, D. L. M., Agnoli, F., & Loftus, E. F. (2007). Changing history: Doctored photographs affect memory for past public events. *Applied Cognitive Psychology, 21*, 1005-1022.

Scoboria, A., Wysman, L., & Otgaar, H. (2012). Credible suggestions affect false autobiographical beliefs. *Memory, 20*, 429-442.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31*, 713-724.

Slater, A., Tiggemann, M., Firth, B., & Hawkins, K. (2012). Reality check: An experimental investigation of the addition of warning labels to fashion magazine images on women's mood and body dissatisfaction. *Journal of Social and Clinical Psychology, 31*, 105-122.

Strange, D., Garry, M., Bernstein, D. M., & Lindsay, D. S. (2011). Photographs cause false memories for the news. *Acta Psychologica*, *136*(1), 90-94.

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

Underwood, J., & Pezdek, K. (1998). Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review, 5*, 449-453.

van de Veer, E., de Lange, M. A., van der Haar, E., & Karremans, J. C. (2012). Feelings of safety: Ironic consequences of police patrolling. *Journal of Applied Social Psychology, 42*, 3114-3125.

Wade, K. A., Garry, M., Nash, R. A., & Harper, D. N. (2010). Anchoring effects in the development of false childhood memories. *Psychonomic Bulletin & Review*, *17*, 66-72.

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.

Wade, K. A., Green, S. L., & Nash, R. A. (2010). Can fabricated evidence induce false eyewitness testimony? *Applied Cognitive Psychology, 24*, 899-908.

Winkielman, P., Schwarz, N., Reber, R., & Fazendeiro, T. A. (2003). Cognitive and affective consequences of visual fluency: When seeing is easy on the mind. In L. M. Scott, & R. Batra (Eds.), *Persuasive imagery: A consumer response perspective* (pp. 75-90). Mahwah, NJ: Lawrence Erlbaum.

Wright, D. S., Nash, R. A., & Wade, K. A. (2015). Encouraging eyewitnesses to falsely corroborate accusations: Effects of rapport-building and incriminating evidence. *Psychology, Crime & Law, 21,* 648-660.

Table 1. *Verbatim wording of target questions and scale anchors in Experiments 1-3.*

Question	Experiment	Wording
Protesters	1, 2, 3	Of those people who lined the streets of [the UK / London] to attend the [Olympic torch relay / Royal Wedding], how many per one thousand would you estimate were there to protest rather than to support?
Arrests	1, 2, 3	Of those people who lined the streets of [the UK / London] to attend the [Olympic torch relay / Royal Wedding], how many in total would you estimate were arrested at the scene?
Injuries	1	Of those people who lined the streets of the UK to attend the Olympic torch relay, how many in total would you estimate were injured at the scene?
Police officers	2	Approximately how many extra police officers were on call in Central London on the day of the event to ensure a safe and secure environment?
Violence	1, 2, 3	How violent were those people who protested on the streets at the [Olympic torch relay / Royal Wedding]? [Not at all violent; Very violent]
Damage	1	How much damage to local property do you think was caused by spectators at the Olympic torch relay? [Very little damage; A very great deal of damage]
Smoothly	1	Do you think that the overall Olympic torch relay went as smoothly as expected? [Not at all smoothly; Very smoothly]
Police control	1, 3	How would you rate the police's success in controlling the [Olympic torch relay crowds across the country / Royal Wedding crowds]? [Highly unsuccessful; Highly successful]
Supportive-self	2, 3	How supportive were you personally of the Royal Wedding? [Not at all supportive; Very supportive]
Supportive-public	2, 3	How supportive were the British Public of the Royal Wedding? [Not at all supportive; Very supportive]
Positive-self	2, 3	How positive or negative did you personally feel the Royal Wedding was? [Very negative; Very positive]
Positive-public	2, 3	How positive or negative do you think the British Public in general felt the Royal Wedding was? [Very negative; Very positive]
Attend in future	1, 2, 3	How likely are you to want to attend an event similar to the [Olympic torch relay / Royal Wedding] if it were to happen in the UK again in future? [Very unlikely; Very likely]
Protest in future	1, 2, 3	How likely are you to protest an event similar to the [Olympic torch relay / Royal Wedding] if it were to happen in the UK again in future? [Very unlikely; Very likely]

Table 2. *Responses to target questions in Experiment 1. The left side of the table reports median responses for each condition, alongside means and standard deviations (in parentheses) for illustrative purposes. The right side of the table reports nonparametric omnibus tests and, where these tests are statistically significant, pairwise contrasts of conditions.*

Question	Condition			Omnibus test	Test of pairwise contrasts		
				Kruskal-Wallis χ^2	Dunn *z*		
	Genuine (A)	Doctored (B)	Doctored-disclaimer (C)		(B) vs. (A)	(C) vs. (A)	(C) vs. (B)
Protesters (# per 1000)	*Mdn* = 5 *M* = 33.00 (125.32)	*Mdn* = 35 *M* = 67.58 (101.38)	*Mdn* = 10 *M* = 59.07 (103.63)	18.90*** η^2 = .09	4.29***	2.78*	-1.60
Arrests (#)	*Mdn* = 10 *M* = 58.49 (143.12)	*Mdn* = 20 *M* = 115.13 (225.03)	*Mdn* = 20 *M* = 114.49 (296.10)	4.99 η^2 = .02			
Violence (1-5)	*Mdn* = 2 *M* = 1.85 (0.84)	*Mdn* = 2 *M* = 2.08 (0.78)	*Mdn* = 2 *M* = 2.01 (0.90)	3.05 η^2 = .02			
Damage (1-5)	*Mdn* = 2 *M* = 1.94 (0.97)	*Mdn* = 2 *M* = 2.02 (0.86)	*Mdn* = 2 *M* = 1.99 (0.95)	0.56 η^2 = .00			
Police control (1-5)	*Mdn* = 4 *M* = 4.19 (0.63)	*Mdn* = 4 *M* = 3.85 (0.80)	*Mdn* = 4 *M* = 3.92 (0.73)	7.99* η^2 = .04	-2.63**	-2.22*	0.47
				Mood's Median test χ^2	Mood's Median test χ^2		
	Genuine (A)	Doctored (B)	Doctored-disclaimer (C)		(B) vs. (A)	(C) vs. (A)	(C) vs. (B)
Injuries (#)	*Mdn* = 20 *M* = 169.12 (648.98)	*Mdn* = 20 *M* = 42.52 (57.04)	*Mdn* = 20 *M* = 113.96 (293.45)	1.63 *V* = .09			
Smoothly (1-5)	*Mdn* = 4 *M* = 4.00 (0.76)	*Mdn* = 4 *M* = 3.54 (0.87)	*Mdn* = 4 *M* = 3.66 (0.77)	6.96* *V* = .19	-4.73*	-4.62*	0.01
Attend in future (1-5)	*Mdn* = 4 *M* = 3.63 (1.22)	*Mdn* = 4 *M* = 3.40 (1.47)	*Mdn* = 4 *M* = 3.70 (1.13)	0.52 *V* = .05			
Protest in future (1-5)	*Mdn* = 1 *M* = 1.13 (0.42)	*Mdn* = 1 *M* = 1.34 (0.85)	*Mdn*= 1 *M* = 1.28 (0.73)	2.52 *V* = .11			

* p< .05, ** p< .01, *** p< .001

Table 3. *Responses to target questions in Experiment 2. The left side of the table reports median responses for each condition, alongside means and standard deviations (in parentheses) for illustrative purposes. The right side of the table reports nonparametric omnibus tests and, where these tests are statistically significant, pairwise contrasts of conditions.*

Question	Condition			Kruskal-Wallis χ^2	Dunn *z*		
	Genuine (A)	Well-doctored (B)	Badly-doctored (C)		(B) vs. (A)	(C) vs. (A)	(C) vs. (B)
Protesters (# per 1000)	*Mdn* = 3 *M* = 17.35 (35.96)	*Mdn* = 11 *M* = 62.03 (142.02)	*Mdn* =10 *M* = 45.76 (112.39)	14.56*** η^2 = .08	3.45***	3.16**	-0.31
Arrests (#)	*Mdn* = 20 *M* = 155.75 (670.66)	*Mdn* = 35 *M* = 213.02 (680.73)	*Mdn* = 50 *M* = 213.47 (408.65)	6.77* η^2 = .04	1.63	2.57*	0.94
Police officers (#)	*Mdn* = 2000 *M* = 16828.07 (71329.06)	*Mdn* = 2000 *M* = 8870.00 (19776.87)	*Mdn* = 2000 *M* = 11606.05 (31507.42)	0.14 η^2 = .00			
Violence (1-5)	*Mdn* = 2 *M* = 1.81 (0.92)	*Mdn* = 2 *M* = 2.09 (0.92)	*Mdn* = 2 *M* = 2.05 (0.82)	4.56 η^2 = .03			
Supportive-public (1-5)	*Mdn* = 5 *M* = 4.58 (0.60)	*Mdn* = 4 *M* = 4.33 (0.69)	*Mdn* = 4 *M* = 4.29 (0.59)	8.07* η^2 = .05	-2.11*	-2.71**	-0.60
Supportive-you (1-5)	*Mdn* = 4 *M* = 3.81 (1.20)	*Mdn* = 4 *M* = 3.78 (1.22)	*Mdn* = 3 *M* = 3.59 (1.05)	2.21 η^2 = .01			
Positive-public (1-5)	*Mdn* = 5 *M* = 4.51 (0.66)	*Mdn* = 4 *M* = 4.36 (0.61)	*Mdn* = 4 *M* = 4.24 (0.68)	3.87 η^2 = .02			
Positive-you (1-5)	*Mdn* = 4 *M* = 4.19 (0.95)	*Mdn* = 4 *M* = 4.12 (0.88)	*Mdn* = 4 *M* = 3.92 (0.90)	5.56 η^2 = .03			
Attend in future (1-5)	*Mdn* = 2 *M* = 2.54 (1.36)	*Mdn* = 2 *M* = 2.31 (1.17)	*Mdn* = 2 *M* = 2.29 (1.38)	1.31 η^2 = .00			
	Genuine (A)	Doctored (B)	Doctored-disclaimer (C)	Mood's Median test χ^2			
Protest in future (1-5)	*Mdn* = 1 *M* = 1.16 (0.62)	*Mdn* = 1 *M* = 1.14 (0.40)	*Mdn* = 1 *M* = 1.22 (0.64)	4.52 *V* = .16			

* *p*< .05, ** *p*< .01, *** *p*< .001

Table 4. *Responses to target questions in Experiment 3. The left side of the table reports median responses for each condition, alongside means and standard deviations (in parentheses) for illustrative purposes. The right side of the table reports nonparametric omnibus tests and, where these tests are statistically significant, pairwise contrasts of conditions.*

Question	Condition			Kruskal-Wallis χ^2	Dunn *z*		
	Genuine (A)	Well-doctored (B)	Badly-doctored (C)		(B) vs. (A)	(C) vs. (A)	(C) vs. (B)
Protesters (# per 1000)	*Mdn* = 10 *M* = 28.50 (105.58)	*Mdn* = 20 *M* = 77.97 (157.92)	*Mdn* = 10 *M* = 59.79 (141.55)	14.23*** η^2 = .07	3.74***	2.21*	-1.34
Arrests (#)	*Mdn* = 10 *M* = 34.99 (76.51)	*Mdn* = 15.5 *M* = 44.64 (75.28)	*Mdn* = 20 *M* = 70.44 (253.99)	6.30* η^2 = .03	2.03*	2.27*	0.34
Violence (1-7)	*Mdn* = 2 *M* = 2.31 (1.07)	*Mdn* = 2 *M* = 2.64 (1.24)	*Mdn* = 2 *M* = 2.26 (0.99)	3.73 η^2 = .02			
Police control (1-7)	*Mdn* = 6 *M* = 5.62 (1.19)	*Mdn* = 6 *M* = 5.64 (1.19)	*Mdn* = 6 *M* = 5.48 (1.28)	0.53 η^2 = .00			
Supportive-public (1-7)	*Mdn* = 6 *M* = 5.72 (1.09)	*Mdn* = 6 *M* = 5.84 (0.99)	*Mdn* = 6 *M* = 5.89 (1.04)	1.11 η^2 = .00			
Supportive-you (1-7)	*Mdn* = 5 *M* = 5.14 (1.72)	*Mdn* = 5.5 *M* = 4.91 (2.10)	*Mdn* = 5 *M* = 5.08 (1.93)	0.18 η^2 = .00			
Positive-public (1-7)	*Mdn* = 6 *M* = 5.61 (1.08)	*Mdn* = 6 *M* = 5.79 (1.02)	*Mdn* = 6 *M* = 5.76 (1.10)	1.31 η^2 = .00			
Positive-you (1-7)	*Mdn* = 6 *M* = 5.36 (1.68)	*Mdn* = 6 *M* = 5.57 (1.74)	*Mdn* = 6 *M* = 5.47 (1.52)	1.12 η^2 = .00			
Attend in future (1-7)	*Mdn* = 3 *M* = 3.23 (2.06)	*Mdn* = 3 *M* = 3.24 (1.91)	*Mdn* = 4 *M* = 3.63 (2.04)	1.72 η^2 = .00			
				Mood's Median test χ^2			
	Genuine (A)	Doctored (B)	Doctored-disclaimer (C)				
Protest in future (1-7)	*Mdn* = 1 *M* = 1.42 (1.06)	*Mdn* = 1 *M* = 1.14 (0.63)	*Mdn* = 1 *M* = 1.29 (0.93)	5.11 *V* = .16			

* *p*< .05, *** *p*< .001

Table 5. *Overall effect size* d_A *estimates for each target question across experiments. Where a target question was asked in more than one experiment, effect sized estimates are based on random-effects mini meta-analyses. Figures in the first set of square brackets in each cell represent 95% confidence intervals; figures in the second set of square brackets in each cell represent 99.9% confidence intervals.*

Question	Experiments	Estimate of effect size d_A for pairwise contrast of groups [95% CI] [99.9% CI]		
		Believable fake vs. Genuine	Unbelievable fake vs. Genuine	Unbelievable fake vs. Believable fake
Protesters (# per 1000)	1, 2, 3	**0.70** **[0.50, 0.90]** **[0.36, 1.04]**	**0.50** **[0.30, 0.70]** **[0.16, 0.84]**	-0.20 [-0.40, 0.00] [-0.53, 0.14]
Arrests (#)	1, 2, 3	**0.32** **[0.12, 0.52]** [-0.01, 0.66]	**0.43** **[0.23, 0.63]** **[0.09, 0.77]**	0.11 [-0.09, 0.31] [-0.22, 0.45]
Injuries (#)	1	0.08 [-0.27, 0.43] [-0.50, 0.66]	0.11 [-0.23, 0.43] [-0.46, 0.67]	0.05 [-0.29, 0.40] [-0.52, 0.63]
Police officers (#)	2	0.00 [-0.37, 0.37] [-0.61, 0.61]	0.07 [-0.29, 0.44] [-0.54, 0.69]	0.05 [-0.31, 0.42] [-0.56, 0.66]
Violence (Likert)	1, 2, 3	**0.29** **[0.09, 0.50]** [-0.05, 0.64]	0.17 [-0.05, 0.38] [-0.20, 0.53]	-0.14 [-0.34, 0.06] [-0.47, 0.20]
Damage (Likert)	1	0.13 [-0.22, 0.47] [-0.45, 0.70]	0.06 [-0.28, 0.39] [-0.50, 0.62]	-0.06 [-0.40, 0.27] [-0.63, 0.50]
Smoothly (Likert)	1	**-0.52** **[-0.87, -0.18]** [-1.11, 0.06]	**-0.40** **[-0.74, -0.06]** [-0.97, 0.17]	0.13 [-0.22, 0.47] [-0.45, 0.70]
Police control (Likert)	1, 3	-0.19 [-0.65, 0.26] [-0.96, 0.57]	-0.22 [-0.49, 0.06] [-0.67, 0.24]	-0.02 [-0.26, 0.22] [-0.42, 0.38]

Supportive-self (Likert)	2, 3	-0.04 [-0.39, 0.34] [-0.44, 0.37]	-0.12 [-0.40, -0.17] [-0.60, 0.36]	-0.07 [-0.35, 0.22] [-0.54, 0.41]
Supportive-public (Likert)	2, 3	-0.13 [-0.60, 0.34] [-0.92, 0.66]	-0.17 [-0.83, 0.50] [-1.29, 0.95]	-0.01 [-0.26, 0.23] [-0.43, 0.40]
Positive-self (Likert)	2, 3	0.02 [-0.23, 0.28] [-0.41, 0.45]	-0.16 [0.52, 0.20] [-0.77, 0.45]	-0.20 [-0.45, 0.05] [-0.62, 0.21]
Positive-public (Likert)	2, 3	-0.06 [-0.46, 0.35] [-0.74, 0.63]	-0.12 [-0.70, 0.46] [-1.10, 0.85]	-0.07 [-0.32, 0.18] [-0.48, 0.34]
Attend in future (Likert)	1, 2, 3	-0.08 [-0.28, 0.12] [-0.41, 0.25]	0.02 [-0.21, 0.25] [-0.37, 0.41]	0.10 [-0.10, 0.30] [-0.23, 0.43]
Protest in future (Likert)	1, 2, 3	-0.01 [-0.22, 0.21] [-0.37, 0.35]	0.11 [-0.09, 0.30] [-0.23, 0.44]	0.12 [-0.08, 0.31] [-0.22, 0.45]

Note: Effect size estimates are calculated using the *A* method described by Ruscio (2008), and converted to d_A using the procedure described by Li (2016). Confidence intervals are calculated using ESCI software (Cumming, 2012).

Note: 'Believable fake' here corresponds to the 'Doctored' condition of Experiment 1 and the 'Well-doctored' conditions of Experiments 2-3. 'Unbelievable fake' corresponds to the 'Doctored-disclaimer' condition of Experiment 1 and the 'Badly-doctored' conditions of Experiments 2-3.

Note: Confidence intervals that exclude zero are presented in bold

Table S1. *Sample sizes and effect size estimates* (d_A) *for each pairwise contrast of conditions across target questions in each experiment.*

Question	Experiment	Group *n*			Effect size d_A for pairwise contrast		
		Genuine (A)	Believable fake (B)	Unbelievable fake (C)	(B) vs. (A)	(C) vs. (A)	(C) vs. (B)
Protesters (# per 1000)	Experiment 1	66	64	71	0.8187	0.4758	-0.2699
	Experiment 2	57	58	59	0.6954	0.6727	-0.0842
	Experiment 3	74	76	62	0.6049	0.3807	-0.2192
Arrests (#)	Experiment 1	65	62	70	0.3087	0.3640	0.0929
	Experiment 2	57	58	59	0.3310	0.5330	0.1870
	Experiment 3	74	76	62	0.3270	0.3999	0.0714
Injuries (#)	Experiment 1	66	62	70	0.0813	0.1054	0.0544
Police Officers (#)	Experiment 2	57	58	59	0.0000	0.0736	0.0540
Violence (Likert)	Experiment 1	67	65	71	0.2940	0.1640	-0.1150
	Experiment 2	57	58	59	0.3503	0.3643	0.0022
	Experiment 3	74	76	62	0.2411	-0.0282	-0.2759
Damage (Likert)	Experiment 1	67	65	71	0.1261	0.0572	-0.0626
Smoothly (Likert)	Experiment 1	67	65	71	-0.5233	-0.4023	0.1255
Police control (Likert)	Experiment 1	67	65	71	-0.4302	-0.3525	0.0794
	Experiment 3	74	76	62	0.0326	-0.0762	-0.1233
Supportive-self (Likert)	Experiment 2	57	58	59	-0.0245	-0.2730	-0.2243
	Experiment 3	74	76	62	-0.0446	0.0201	0.0673
Supportive-public (Likert)	Experiment 2	57	58	59	-0.3785	-0.5102	-0.0973
	Experiment 3	74	76	62	0.1029	0.1704	0.0611
Positive-self (Likert)	Experiment 2	57	58	59	-0.1245	-0.3544	-0.2667
	Experiment 3	74	76	62	0.1384	0.0154	-0.1443
Positive-public (Likert)	Experiment 2	57	58	59	-0.2718	-0.4245	-0.1723
	Experiment 3	74	76	62	0.1423	0.1686	0.0165
Attend in future (Likert)	Experiment 1	67	65	71	-0.1227	0.0503	0.1601
	Experiment 2	57	58	59	-0.1594	-0.2127	-0.0699
	Experiment 3	74	76	62	0.0241	0.2004	0.1870
Protest in future (Likert)	Experiment 1	67	65	71	0.1508	0.1690	0.0226
	Experiment 2	57	58	59	0.0566	0.2478	0.1994

Experiment 3	74	76	62	-0.2045	-0.0812	0.1365

Note: Effect size estimates are calculated using the *A* method described by Ruscio (2008), and converted to d_A using the procedure described by Li (2016)

Figure S1. Schematic representation of the photograph used in the Doctored-disclaimer condition of Experiment 1, demonstrating the position and prominence of the disclaimer (original image blurred).