Comparing eye gaze tracking to reported perceptions of manipulated and unmanipulated digital images

Sabrina B. Caldwell, Tamás D. Gedeon, Richard L. Jones, Martin Henschke
Research School of Computer Science, The Australian National University
Acton ACT 0200 Australia

Abstract

To investigate human perceptions of image manipulation at both the conscious and non-conscious levels, we compared participants' verbal reporting of image manipulation to data recordings of their eye movements while viewing 36 images of varying manipulation levels. To further understand subjects' ability to use image comparison tools to aid in manipulation detection, variants of a trial 'image packaging' software provided two levels of image comparison support tools.

Keywords: Eye gaze; manipulated images; detect image manipulation

Introduction

Increasingly, we encounter our information about the world in image form. At the same time, the ability of humans to manipulate images is greater than at any time previously in history. While there is research on the use of manipulated images in advertising[1] there is very little understanding of the effects of ubiquitous photo manipulation such as is rife in social media and family photos.

As a step in understanding this phenomenon, this experiment uses both eye gaze tracking and verbal questioning to compare what subjects see (as represented by their eye gaze tracking results) and what they perceive (as represented by their question responses) when provided with both standalone images and images that have been packaged with additional assistive information.

This experiment investigates peoples' ability to see manipulations in images, and seeks to identify whether providing additional comparison images along with the presentation image enables participants to identify manipulations in images more accurately and/or quickly.

In addition, the experiment attempts to determine how subjects interpret images in relation to any manipulations they contain.

[1] For example effect of airbrushed models on teen body image [Grabe et al 2008]

Materials and Methods

Twelve volunteer participants undertook eye gaze tracking and verbal questioning as they viewed images of photographs ranging from unmanipulated to strongly manipulated. The participants' mean age was 32.7 (SE 11.1) years.

Facelab 5.0.2 by Seeing Machines was used to track eye gaze with two infra-red (IR) cameras and a single IR light emitter pod centrally located in front of and below the monitor displaying the images. Eyeworks v3.8, also by Seeing Machines, was used for experiment delivery, recording and analysis.

Subjects were shown three sets of 12 images each comprising 3 unmanipulated images and 9 images manipulated by splicing in or erasing elements of varying sizes from the scene (examples Figure 2). The first set comprised standalone images. The second and third sets were presented in a mobile, self-contained image format (MSCI), a trial 'image packaging' software currently under development. The second set presented bundled images in which a presentation image was accompanied by the original image for comparison. This configuration was repeated in the third set but also accompanied by a 'differences map' image that highlighted any changed pixels.

Images within sets were varied in order using Latin square randomisation to avoid any ordering bias. In each subsequent set some images were repeated in the new format to identify how much assistance subjects needed to identify manipulations.

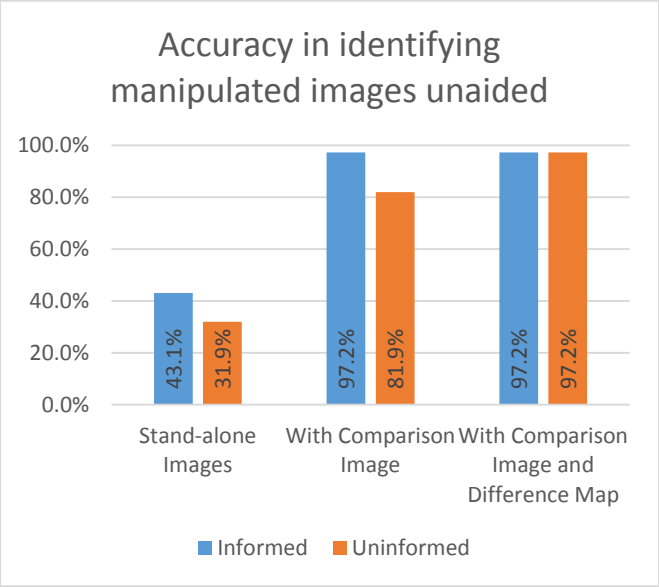


Figure 1: Image manipulation identification accuracy n=12

Subjects were assigned to one of two sub-groups, those who had pre-existing familiarity with image manipulation issues through exposure to the authors' research (informed), and those who did not (uninformed). In some cases (4 subjects) an additional set of eight images were employed after the common part of the experiment, to further test whether focussed exposure to image manipulation predisposes subjects to identify manipulated images with increased accuracy.

At the same time as their eye gaze was being tracked, participants were asked a short set of questions relevant to each phase of the experiment and their responses recorded. These questions targeted their perception of any manipulations that might appear in the images they viewed, and their interpretations of the images.

Finally, participants completed a short survey and responded to open-ended questions about their attitudes towards image manipulation.

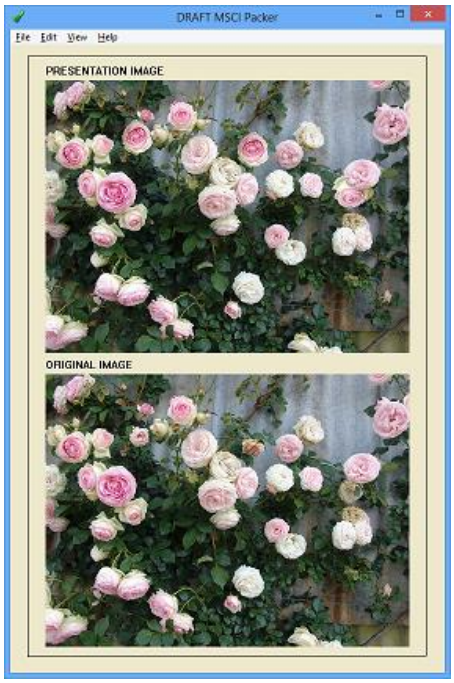
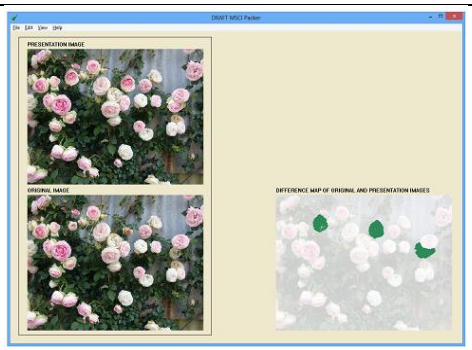
Set	Roses copy/move	Pier retouch
Set 1- Standalone image		
Set 2 – Comparison original		
Set 3 – Comparison original and difference map		

Figure 2: Examples of images shown to subjects in Experiment A

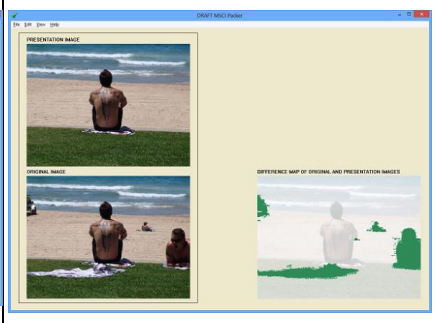
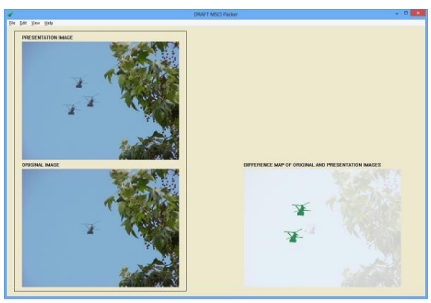
Coins copy/move	Man on beach retouch	Helicopter copy/move
Not shown in set 1.	*Not shown in set 1.*	*Not shown in set 1.*
		*Not shown in set 2.*
		

Figure 2 (continued)

Results

Overall, the ability of participants to verbally identify manipulated images with no assistance was weak, only 37.5% accuracy on average (Figure 1). This was despite eye gaze data indicating that subjects had looked directly at the manipulated areas with greater intensity than would be predicted by the area of the manipulated regions of the image (Table 1). Informed subjects were more likely to report image manipulations than uniformed subjects (43.1% vs 31.9%).

Regions of manipulation: % of participant views in relation to area of manipulated regions in Set 1									
	Roses	Wren	Mosaic	Pier	Girls	Shark	Howard	Milkman	Missile
% of gaze used in manipulated region(s)	12.7%	4.4%	5.7%	5.7%	15.6%	26.0%	27.2%	72.1%	36.7%
Area of manipulated region(s) in pixels	37863	4247	3079	2958	73494	6534	43617	95715	12665
Area of image in pixels	595337	594520	597908	593292	593025	159944	595856	599404	91921
Manipulated region % area of image	6.4%	0.7%	0.5%	0.5%	12.4%	4.1%	7.3%	16.0%	13.8%
% of views in relation to % of area	200%	611%	1116%	1151%	125%	636%	372%	451%	266%

Table 1: Comparing ratio of participant views to area size of manipulated regions

When given a comparison original image, participants' success rate at identifying manipulations more than doubled to 85.4%, although often they either could not say what had been changed or else misreported or under reported what had been changed. Eye gaze data indicated that participants gaze traversed the two images to identify and compare areas of difference. Again, there was a difference in accuracy between informed and uninformed subjects (97.2% vs 81.9%).

When also given a difference map in set 3 of the experiment, participants' success rate in identifying manipulated images increased to 97.2% for both groups, and the difference in quality of perception of manipulations increased. Eye gaze data indicated that participants' gaze referred to the difference map as an aid in locating manipulated areas presentation (manipulated) images as compared with the original.

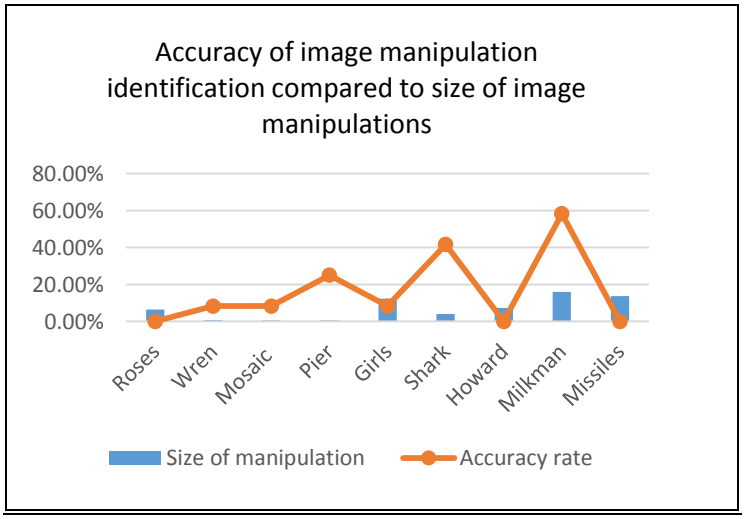


Figure 3: Image manipulation size to identification success

Using the difference map, participants could identify 90.1% of specific manipulations in

detail. When offered a comparison map but not a difference map, participants accurately identified the specific manipulations only 68.1% of the time. Further, incorrect alternative explanations for the effects of manipulations were given, for example the insertion of three additional roses into an image was described by three participants as increased colour contrast.

Moving from the second set to the third set also increased speed of identification of manipulated images, with the time from image appearance to the decision point reducing from 12.4 to 5.3 seconds (standard deviations of mean in seconds were 5.16 and 2.99 respectively). In both cases this compares favourably with the time required for potential identification of manipulated images in Set 1 in which participants were presented with standalone images, 45.7 seconds (SD 7.1).

There was little correlation between the size of the manipulation and the accuracy rate (Figure 3).

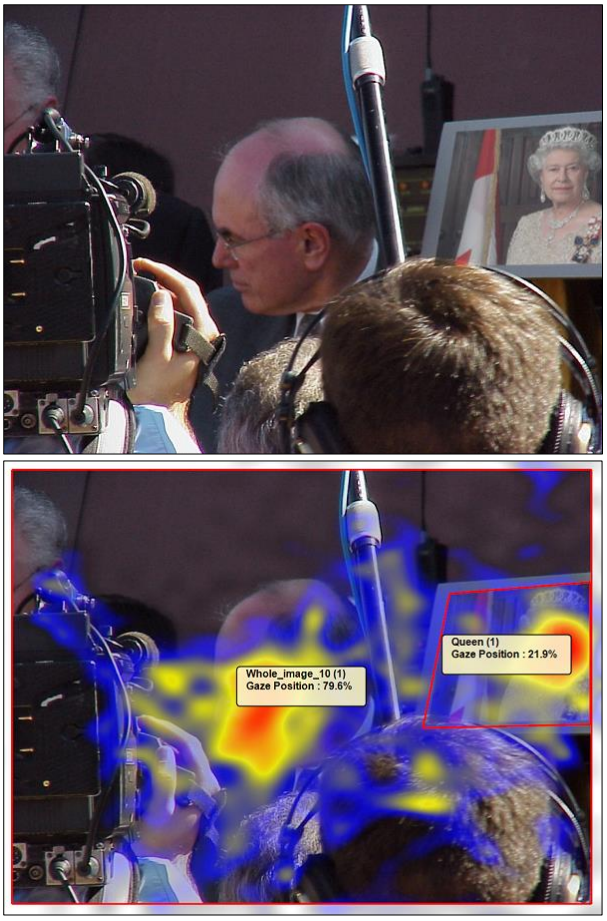


Figure 4: Set 1 standalone image with and without eye gaze 'heat map' annotation.

Comparing eye gaze tracking to question responses indicated that subjects' eye gaze fixated on regions of manipulation in images even when they did not report the image as altered. Overall, while participants only identified 37.5% of the nine manipulated images in Set 1 (standalone images), their eye gaze rested in the regions of manipulation up to 11 times as often as the area the manipulations occupied would predict (Table 1).

As an example, the image of Australian Prime Minister (1996-2007) John Howard with a spliced image of Queen Elizabeth II from Set 1 standalone images was not verbally reported as manipulated by any participant despite the eye gaze 'heat map' demonstrating that the region of the spliced image was viewed intensively at an average 27% of eye gaze within area of the photograph and over 4 times more frequently than would be seen by area alone (Figure 4).

In responding to the survey questions, all participants stated that they cared about photo credibility (average 8.7 out of 10 indicating that on average all the subjects care a lot), societal implications of photo manipulation (average 9.1 out of 10 indicating this significantly matters), and the lack of photo authentication solutions available (average 8.25 out of 10 indicating concerned to very concerned).

Responses to the question "How easy is it for you to tell if a digital photograph you are looking at has been manipulated?" yielded answers clustered around the midway point indicating a middle ground between difficult and easy (average 4.9 out of 10).

In post experiment open-ended questioning, 10 (83%) of participants responded that they use one or more photo editing software packages including Adobe Photoshop, Microsoft Paint, Fireworks, Instagram, Gimp and Adobe Illustrator. The uses to which these software systems are put include cropping, red-eye reduction, colour adjustment, light adjustment, 'fun filters' in the case of Instagram, and making collages.

In response to the question of how they identified manipulations when viewing photographs, the strategies identified by participants were "searching for sharp edges," "when things don't look right, like one face on another person's body," "lighting effects," "if the dimensions are wrong," if things are too perfect," and "shadows going the wrong way." Some participants (25%) stated they did not know how to identify manipulations in images.

When asked what they look for in a photograph, with suggestions of aesthetics, meaning, or representations of reality, participants responded overall that they looked first for aesthetics, then meaning, then representations of reality.

Discussion

Participants brought a diverse range of understanding of photographic images to the viewing exercise. This understanding often informed their detection of image manipulation. Most significantly, participants who were aware that the research involved photo credibility were more successful in identifying manipulations (Figure 1), which may have resulted from them viewing the photographs with a more critical eye (that is, perhaps they were more likely to consciously pay attention to the results of their non-conscious identification of changed regions of the photograph).

Figure 5: Hoax photo of cow on bonnet of BMW sedan

Knowledge and life experiences generally played a role in participants understanding the meaning of the photographs they viewed. For example, of the 12 participants, only 1 articulated the connection between the image of Queen Elizabeth spliced into the photograph of then Prime Minister John Howard in Figure 4 above (John Howard's well known monarchist views on whether Australia should be a Republic). Three subjects were unable to identify John Howard at all and focussed on the Queen or the media aspects of the image.

It was expected that manipulations of larger sizes would be spotted more readily than manipulations of smaller sizes, this was not the case. There was no obvious correlation between the size of the manipulation and the accuracy rate (Figure 3). In an image used in all three sets, in which it appears three people are about to jump over a pier rail next to a sign that reads "JUMPING FROM PIER" (from which the word 'NO' was erased from the image) the overall size of the manipulation was only 0.5% yet the accuracy rating of participants identifying the manipulation was 25%. At the same time, an image of missiles in which 13% of the image was an additional spliced image had a 0% success rate of manipulation identification.

Given the tendency of participants to rationalise elements in images (discussed below), it may be that a more defining characteristic of more easily discernible manipulations is their saliency, i.e. the extent to which the elements added to or removed from an image contribute to the understanding of that image. This suggests that further experimentation to tease out the differences between apparent size and saliency impact may be useful.

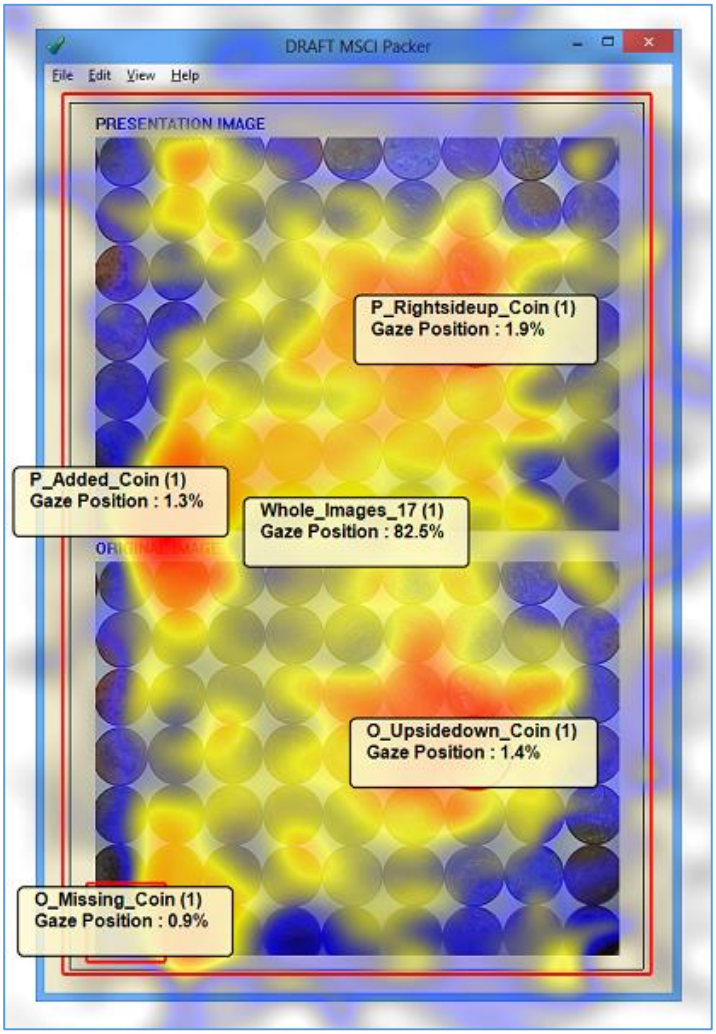


Figure 6: Manipulated coins photo: presentation image provided with comparison original

While all participants stated at a level of 8.25 or over out of 10 that they cared about photo credibility, societal implications of photo manipulation, and the lack of photo authentication solutions, their verbal exposition when discussing the meaning of the photographs presented to them indicated that they were more likely to justify the oddness of the image than to question it.

This was true even when the subject of image manipulation had been discussed moments earlier. Short-term increased awareness of image manipulation issues (as represented by participation in the experimental study) appeared to have little effect in conditioning participants to look at photographs critically. Four participants who had been assigned to the untrained group prior to the experiment participation were also asked to view 8 additional images, 2 of which were unmanipulated and 6 of which were manipulated similarly to the common part of the experiment. Three

responded “no” in response to the query “Do you believe that any of these eight photos has been faked?” The 4th subject responded yes but could not identify more than one faked photograph from the 6 presented.

For example, although common sense would dictate that the photograph in Figure 5 had been manipulated, participants explained it away with justifications such as “maybe the car was warmer to sleep on than the snow,” “relates to the use of leather in cars,” or an inspirational message of unknown origin “don’t think that anything is impossible.”

This matches the uncertainty of responses to the survey question about ways to identify photograph manipulations, as well the verbal exposition and eye gaze data in which participants often used words indicating uneasiness with a photograph as they looked at manipulated elements in images, such as one participant commenting “that shouldn’t be up there” as her eye gaze rested on the cow in the image at Figure 5.

It was noted that a ‘hiding’ effect occurred when additional, less obvious alterations were included in an image. In Figure 6 an image of a field of coins was presented to subjects for the first time in Set 2 of the experiment wherein subjects were offered both a presentation image and the original for comparison. Although most (11 of 12) noted that a coin had been added to the original, only one participant noted that another coin in the image had been rotated 180 degrees. This was despite eye gaze tracking identifying that subjects looked at the rotated coin more (2.3% of area within photo) than the added coin (1.6% within photo).

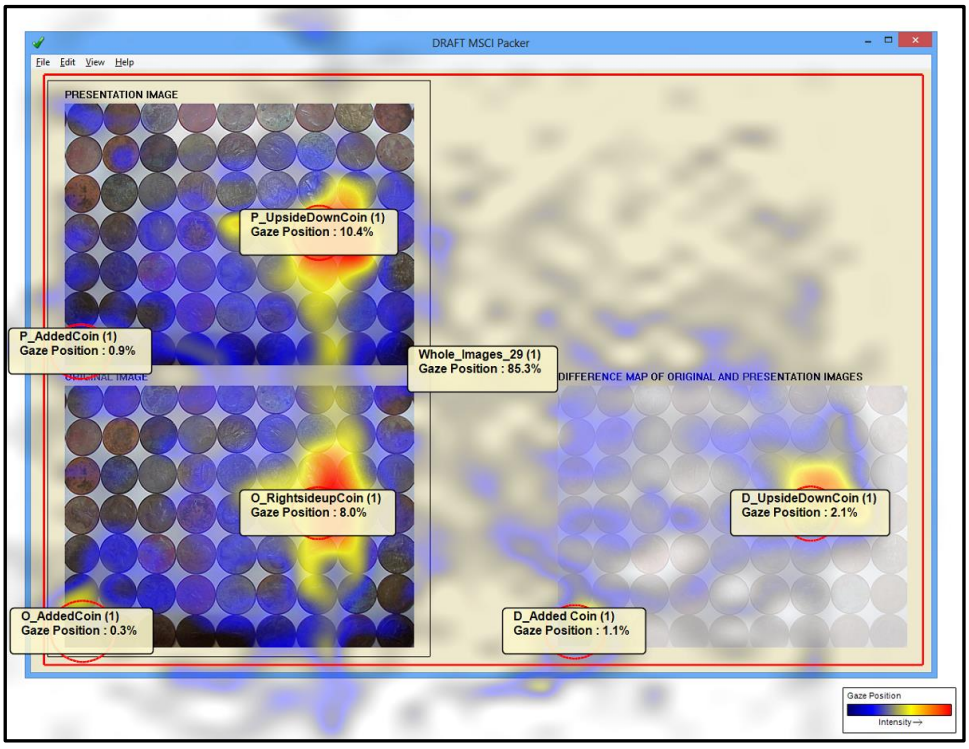


Figure 7: Subjects seek second manipulation when offered a difference map

When the coins photograph was presented again in the third phase of the experiment in which they were offered a difference map identifying changed pixels (Figure 7), all subjects used an eye gaze strategy that compared the pixels demarcated in the difference map as changed to

find and identify the rotated coin in the presentation image by comparison with the original image.

In some cases the use of the difference map in Set 3 enabled participants to confirm previously identified differences noted in Set 2, somewhat like an answer key. In other cases the difference map provided participants with information that enabled them to identify that there were differences or additional differences in images they had previously passed as unchanged. Overall, using the image configuration in Set 3 provided the greatest level of accuracy (97.2% manipulated images identified, 90.1% of all manipulated regions identified) and speed (5.3 seconds on average).

It may be that one reason subjects are more likely to explain than to question manipulated images is that they want to believe they can spot fakes and therefore seek alternative explanations for unlikely elements in the images. Conversely, it may also be the case that subjects feel they can't spot fakes and therefore focus on and rationalise the meanings of the images.

Figure 8: Manipulated sea anemones in frozen pond photograph

These rationalisations can be quite unexpected, as in the case of a participant attempting to explain the juxtaposition of a trio of sea anemones and a frozen pond (Figure 8). In this case the participant focussed on the anemones being sea creatures and reinterpreted the snow on the edging stones as salt.

It is worth noting that the co-existence of high levels of concern about photo manipulation and the tendency to justify rather than identify manipulated images is dichotomous. It may be that photographs have not yet shed their cachet of being representations of reality; subjects are conditioned to look at images as 'real'. This would also be useful to examine in future experiments.

Summary

Comparing eye gaze tracking to question responses reveals that subjects may see more of the changes in manipulated images than they consciously report.

It is not necessarily the case that larger manipulations are more easily seen. The saliency of the manipulation may influence the identification success rate, and this bears further investigation.

It appears that it may be that when an image has an obvious manipulation, other lower profile manipulations may not be consciously identified even when they are viewed by the eye gaze.

Subjects who were aware that this research involves manipulated images (the informed group) looked more critically for possible manipulations and performed better when presented with both standalone image and comparison images than those who were uninformed.

Exposing participants to additional standalone images subsequent to the experiment proper did not result in improved performance comparable to the 'informed' cohort in identifying manipulations.

There is a dichotomy between the high levels of concern expressed about photo manipulation and the tendency of participants to explain away manipulations instead of identifying them.

The two levels of MSCI image bundling assisted subjects in perceiving image manipulations more accurately and quickly.

Attributions

All photographs and photo alterations by Sabrina Caldwell with the following exceptions:

Queen Elizabeth II (photograph) used in John Howard photograph.
Source: http://en.wikipedia.org/wiki/File: Queen_of_canada_wob.jpg. Accessed 17 January 2014. Work is assumed to be copyrighted and used under the fair use provisions of copyright law.

Cow on BMW. Originally tweeted by Surrey Roads Police. Prior origin unknown. Source: www.flickr.com/photos/96057563@NO2/9430346171/. Accessed 17 January 2014. Work is assumed to be copyrighted and used under the fair use provisions of copyright law.

References

Goldberg, Joseph H. and Xerxes Kotval. Computer interface evaluation using eye movements: methods and constructs. International Journal of Industrial Ergonomics 24 (1999) 631-645.

Grabe, S., & Ward, L., & Hyde, J. (2008). The role of the media in body image concerns among women: A meta-analysis of experimental and correlational studies. Psychological Bulletin, 134, 460-476.

Henderson, John M. Human gaze control during real-world scene perception. TRENDS in Cognitive Sciences Vol. 7 No. 11 November 2003

Loftus, Geoffrey R and Norman H. Mackworth. Cognitive determinants of fixation location during picture viewing. Journal of Experimental Psychology: Human Perception and Performance, Vol 4(4), Nov 1978, 565-572. DOI: 10.1037/0096-1523.4.4.565.

Underwood, Geoffrey and Tom Foulsham. Visual saliency and semantic incongruency influence eye movements when inspecting pictures. The Quarterly Journal of Experimental Psychology, 59:11, 1931-1949, DOI: 10.1080/17470210500416342.

Seeing Machines. seeingmachines.com

Yarbus, Alfred L. Eye Movements and Vision. New York: Plenum Press, 1967