Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (EECSS 2015)
Barcelona, Spain – July 13 - 14, 2015
Paper No. 308

Imperfect Understandings: A Grounded Theory And Eye Gaze Investigation Of Human Perceptions Of Manipulated And Unmanipulated Digital Images

Sabrina Caldwell, Tamás Gedeon, Richard Jones, Leana Copeland
The Australian National University, Research School of Computer Science
Acton, ACT, Australia
sabrina.caldwell@anu.edu.au, tom.gedeon@anu.edu.au,
richard.jones@anu.edu.au, leana.copeland@anu.edu.au

Abstract **–**To investigate the extent to which humans are able to perceive manipulated images, and how they interpret these images, we use eye gaze tracking and grounded theory analysis of 80 participants viewing a combination of manipulated and unmanipulated images together with pre-familiarisation with image manipulation techniques. We find that accurate identification of manipulations is poor to moderate, and ability to identify *what* has been manipulated is poor. However, extended visual attention to manipulated regions was associated with greater accuracy. Grounded theory analysis shows that uncertainty in respect of veracity of images viewed is common across all images. Similarly, the use of logic influences participants' success rate in identifying manipulations.

Keywords**:** Eye gaze tracking; manipulated images; detect image manipulation

1. Introduction

Increasingly, we encounter information about the world in image form (Kress, 1996). Users of Facebook alone upload more than 350 million photos every day (Facebook, 2013). At the same time, image manipulation is no longer restricted to specialist photographers, but is achievable by anyone with image editing software. Issues relating to inappropriate image manipulation have been noted in a range of disciplines including news reporting (Wheeler, 2002), medicine (Prasad, 2011) and scientific journals (Cromey, 2010).

There has been significant research on how we view images. Examinations of saliency in features of an image have shown that we focus on salient features of an image to discern its meaning (Itti and Koch, 2000 and Treisman, 1980), and investigations of eye gaze scan paths demonstrate that some characteristics of images, especially luminance (Harding and Bloj, 2010) attract our attention. However, to understand how humans interact with the myriad digital images of unknown veracity presented to us each day, it is also of interest to understand to what extent we can discern 'faked' photos, and how we interpret the meanings presented in them. In previous work the authors found participants' success rate to be related to the availability of reference images (Caldwell et al 2015).

As a step in further understanding aspects of computer-mediated visual communication, this experiment uses both eye gaze tracking and verbal questioning to compare what subjects see, as represented by their eye gaze tracking results, to what they perceive, as represented by grounded theory analysis of their verbal responses.

2. Materials and Methods

Eighty volunteer participants undertook eye gaze tracking and verbal questioning while viewing 14 manipulated and unmanipulated photographic images. Volunteers were sourced largely from a first year

computer course and augmented with others from a range of sources. Participants' mean age was 24.4 (SD 8.7); 53 were male and 27 were female.

Image manipulations used were *copy/move* (one part of an image cloned and repositioned within the photograph) and *splicing* (some or all of a secondary image incorporated into a primary image). An additional image was globally changed to grayscale from colour, with both images presented.

Participants were divided into two cohorts of 40 each. Both cohorts viewed 9 common images comprising 5 manipulated and 4 unmanipulated images. Each cohort viewed 5 additional images, differentiated so that one cohort saw the original unmanipulated versions of the images and the other saw the manipulated versions of these images. The images were chosen to cover a range of topics including social images, nature, and politics (Fig. 1).

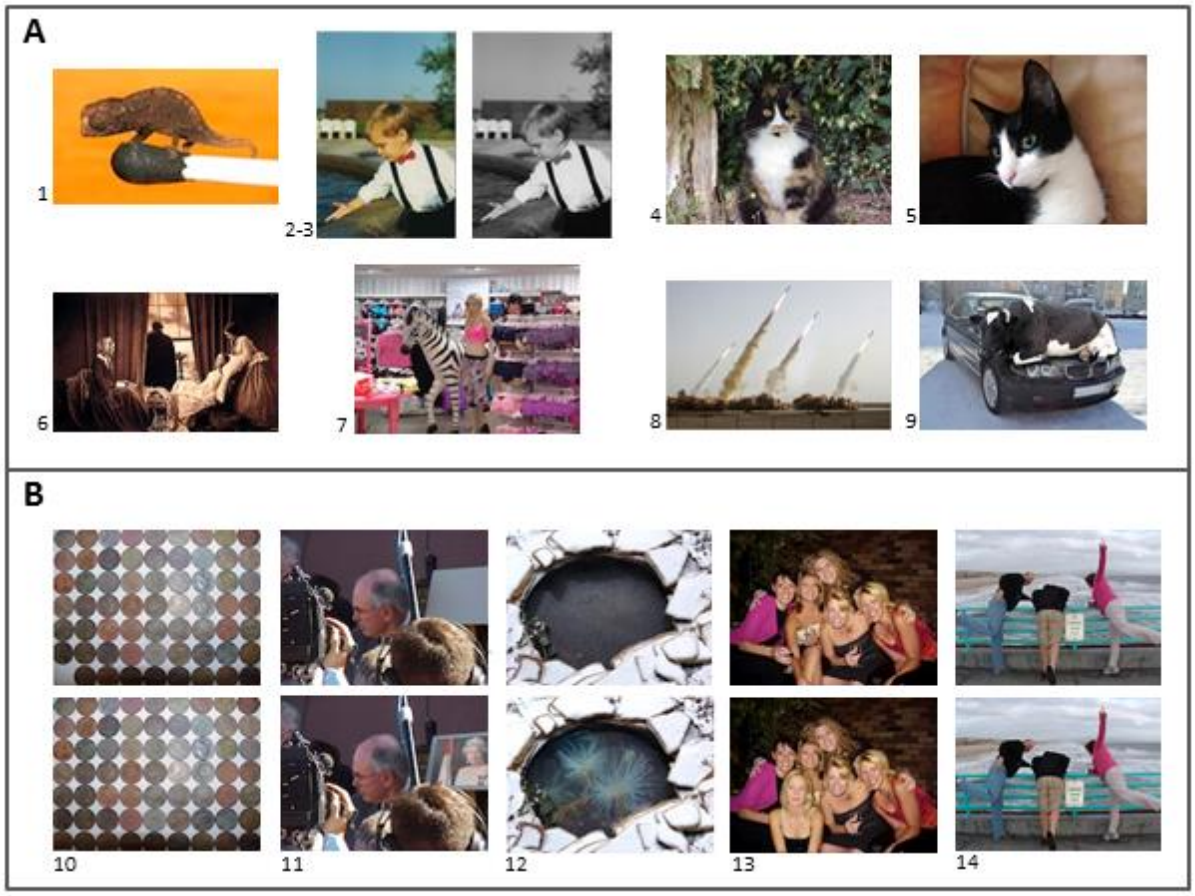


Fig. 1. Common images (A) and images differentiated between cohorts (B)

To support their ability to perceive manipulated features in images, participants were presented with text describing image manipulation techniques prior to viewing images.

For all images, each participant was asked the following questions in order and their freeform answers were recorded:

- What in this image do you find interesting; what attracts your attention?
- Do you believe this image has been manipulated or has not been manipulated? Why?

In two selected cases (images 8 and 13, Figure 1) participants were told the nature of the manipulation (spliced missile and spliced woman respectively) and asked if they could identify which had been added.

The Glaserian grounded theory approach was used to analyse responses (Glaser, 2007).

Participants' eye gaze was recorded using two Facelab 5.0.2 infra-red cameras and a single IR light emitter pod centrally located below the monitor displaying the images. Eyeworks v3.8 was used for experiment design, delivery, recording and analysis, and to record video evidence of each experiment.

Regions of manipulation were transparently demarcated in relevant images and participants' eye gaze fixations were analysed to determine intensity of viewing of these regions.

3. Results

The overall ability of participants to state whether images presented to them were manipulated or not was poor to moderate, with a 56.0% mean accuracy rate based on 600 accurate identifications across a total of 1071 valid image views (reasons for invalidation included eye gaze registration failure and six participants stating they had previously seen an image). Participants had greater success correctly identifying images as unmanipulated (61.3%) than correctly identifying images as manipulated (50.1%).

Comparing these results to participants' eye gaze for the manipulated images (Table 1) shows that the relationship between eye gaze and accuracy varies widely between images with statistical significance of the difference ranging from not at all significant (0.941 for fixations and 0.908 for duration of gaze) to very significant (0.009 for fixations and 0.002 for duration of gaze). However, of the 8 manipulated images, 3 showed a significant difference between the mean fixations and durations of viewing the manipulated regions between those who stated that the images were manipulated and those who stated that they were not. Across all of the images, there was a significant difference ($p < .001$) between the 732 observations of fixation and duration of eye gaze of those correctly identifying images as manipulated and those who incorrectly identified them as unmanipulated. Fixations were: mean 38.4, sd 25.2, n=213 for those who stated image was manipulated, and: mean 27.3 sd 22.3, n=153 for those who stated image was unmanipulated; durations were mean 10.3 sd 7.6, n=213 for those who stated image was manipulated and mean 7.3 sd 6.6, n=153 for those who stated image was unmanipulated. When Images 10, 11, and 12 are omitted, the difference in fixations between these two outcomes remains statistically significant.

The overall ability of participants able to identify *what* was manipulated was considerably lower, with only 132 correct identifications of what was manipulated out of 477 valid views of manipulated images, or 27.7%.

4. Discussion

Despite extensive pre-familiarisation with concepts of digital image manipulation, participants were poor to moderate at identifying whether images were manipulated or not. However, we note that the average accuracy rate of participants in this experiment (56.0%) compares favourably to the authors' results in a previous experiment (Caldwell et al, 2015) in which participants with only a brief familiarisation with principles of image manipulation achieved a lower mean accuracy rate of 37.5%.

Even when participants were told directly in the case of Images 8 and 13 that a feature had been added (missile and girl added), the ability to then identify which missile (missile number 3) and which girl (the girl in front) had been added was poor. For example, after participants of the cohort viewing the manipulated images were told an additional girl had been spliced into Image 13, their accuracy of what had been manipulated increased from 11 out of 40 participants to 28 out of 40 participants, but 11 participants nominated other girls in the image as spliced in as well.

4. 1. Grounded Theory

We consider grounded theory outcomes at two levels: individual images and overall.

At an individual level, the characteristics of each image (semantics and content) yielded varying outcomes aligned with the nature of the image itself. Elements of interest to participants differed for each image, and corresponding determinations of what had been changed in images they stated had been manipulated, were also image-centric.

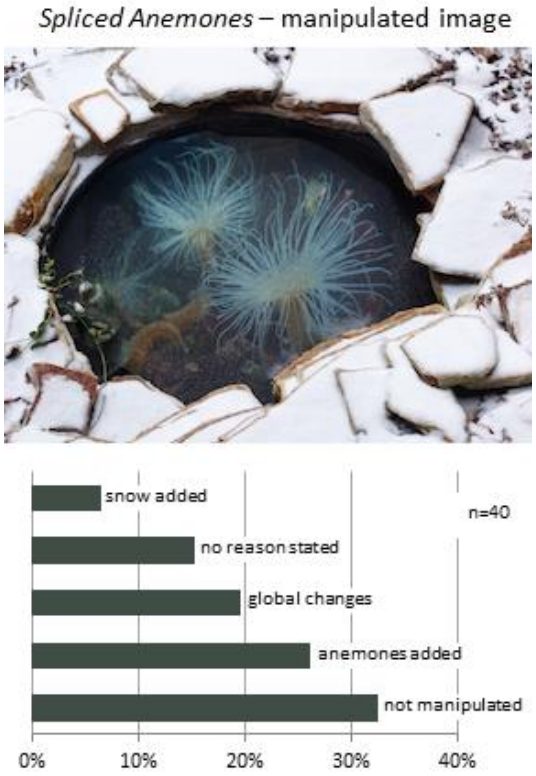


Fig. 2. 67% of participants accurately identified this image as manipulated. 26% accurately identified that the anemones were added

Table. 1. Difference of participants' eye gaze viewing manipulated areas of images and their determinations of manipulation[1]

Image	n (total)	Stated: manipulated (correct)					Stated: unmanipulated (incorrect)					Sig (p<.05)	
		Mean Fixation	SD	Mean ∑ Duration (secs)	SD	n	Mean Fixation	SD	Mean ∑ Duration (secs)	SD	n	Fixation	Duration
3	63	63.4	24.43	15.39	8.29	42	56.7	25.64	15.14	8.32	21	0.329	0.908
8	67	27.0	18.07	7.66	6.08	39	28.6	23.10	8.11	6.43	28	0.765	0.772
9	67	30.2	20.72	9.60	7.56	54	29.0	17.42	8.76	6.45	13	0.837	0.689
10	20	9.3	4.35	7.75	6.68	7	3.1	2.42	2.32	2.49	13	**0.009**	**0.002**
11	37	38.5	19.09	8.34	4.77	21	25.1	12.03	5.45	3.67	16	**0.014**	**0.045**
12	35	54.5	18.82	13.02	6.65	25	32.1	15.21	6.91	5.11	10	**0.001**	**0.008**
13	34	30.1	19.32	8.24	5.97	13	30.6	18.17	8.69	5.92	21	0.941	0.836
14	36	17.0	9.40	7.54	7.93	12	12.3	6.64	4.28	2.79	24	0.143	0.192

[1] Note that manipulated Image 6 was omitted from Table 1 because although it is known that the image was composited from five separate photographic negatives (Robinson, 1860), the exact composition remains unknown and therefore it was not possible to accurately specify manipulated areas.

4. 1. Grounded Theory

We consider grounded theory outcomes at two levels: individual images and overall.

At an individual level, the characteristics of each image (semantics and content) yielded varying outcomes aligned with the nature of the image itself. Elements of interest to participants differed for each image, and corresponding determinations of what had been changed in images they stated had been manipulated, were also image-centric.

For example, in Image 12 (Figure 2) in which sea anemones were spliced into an image of a snow-covered pond, descriptions of manipulations focused on the anemones, the snow, and global changes such as sharpening and colour filtering. By contrast, in Image 11 in which an image of Queen Elizabeth II was spliced into a media scrum near Australian Prime Minister John Howard, manipulation descriptions focused on the Queen and John Howard. For participants identifying that an image was manipulated, the main descriptions of what they perceived as being changed are listed in Table 2 below.

Table. 2. What participants stating image was manipulated perceived as being manipulated

Image		n	Unmanipulated version	Manipulated version
Common to both cohorts	1	50	chameleon added/resized (23), background changed (19), match enlarged (4)	not presented
	2	21	colourised from black & white (10), global changes (7)	not presented
	3	45	not presented	made grayscale (32), other global changes (4)
	4	19	resized (5), cropped/zoomed (3)	not presented
	5	7	global changes (3)	not presented
	6	40	not presented	sepia/brown filter (22), man at window added (3)
	7	23	zebra added (11)	not presented
	8	40	not presented	missile(s) added (15), smoke added/changed (9)
	9	61	not presented	cow added (49), license plate blurred (9)
Differentiated	10	18	global changes (4)	global changes (6), cropped (2)
	11	29	global changes (3)	Queen added (10), global changes (3)
	12	47	global changes (13)	anemone added (12), global changes (9), snow added (3)
	13	30	global changes (6), red-eye reduction (5)	girl added (11), red-eye reduction (2)
	14	22	sign changed (6), woman added (4)	sign changed (13), woman added (4)

Some themes that arose in individual image descriptions were over-arching across all images presented. In particular, the themes of uncertainty and logic presented themselves in all images. These two aspects of participants' responses are detailed in Table 3 below in which the number of instances of uncertainty and use of logic is shown. A response was flagged as using logic only if the logic process was evident in a participant's verbal response. For example, many participants stated that the chameleon in Image 1 was resized to a tiny size, by which it could be inferred that logically they believed that it was not possible for the chameleon to be that small, however only those who stated that logic aloud were counted. Similarly, while a long delay in responding might imply uncertainty, only those who stated they did not know or used other words of uncertainty in their responses were counted.

Table 3: Participants expressing overall themes of Uncertainty and Logic

Image	n	Uncertainty	Logic
1	80	24	14
2	80	12	5
3	80	13	5
4	80	14	3
5	80	10	3
6	80	11	10
7	80	5	3
8	80	20	12
9	80	22	16
10	40	14	3
11	40	16	9
12	40	13	10
13	40	9	5
14	40	12	11

Participants often expressed their uncertainty by stating that they did not know if an image was manipulated, and making statements of general uneasiness with the image, evidenced by words such as 'perhaps, maybe' and phrases like 'I think,' 'I believe,' 'not sure,' 'doesn't look right.'. Uncertainty was also expressed in overt guessing.

Participants often employed logic (rightly or wrongly) in determining the validity of an image. For example, Image 4 was often (26%) noted as resized because it was pixelated rather than understood to be a low resolution photograph. In Image 13 (manipulated version) the girl was identified as added by 5 participants because she wasn't smiling like the others or because her lighting seemed different. In Image 1 the chameleon was stated as added because 'lizards aren't that small,' and in Image 7 the zebra was noted as added because 'zebras don't belong in a lingerie shop.'

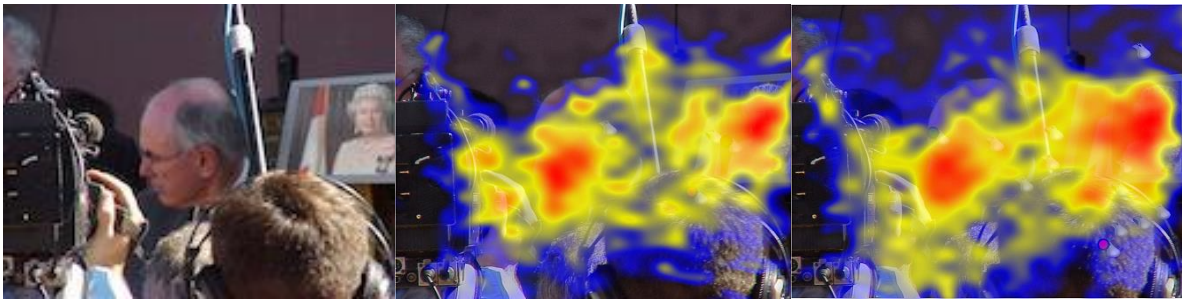
The logic participants used in determining manipulations was informed by their pre-existing knowledge and beliefs. For example, for Image 9, a participant with a farming background commented "I know how much a cow weighs and a car would not stand up to that weight." In the case of Image 11, only 26 out of 80 participants seemed to know that the central figure being interviewed was former Prime Minister John Howard; with others referring to him as 'the man in the middle,' or 'the bald-headed man.' This may have been a function of the relative youth of the participants and the fact that John Howard has been out of office for over 7 years; the average age of those who identified the ex-Prime Minister was 27.7 (SD 11.9) in comparison to experiment average of 24.4 (SD 8.7) years. In the case where a colour image (Image 2) was filtered to create a black & white version (Image 3) participants with photo editing knowledge correctly identified it as manipulated with the use of appropriate language (desaturated, filtered) to describe the change.

However, the use of logic was counterproductive if the logic was not sound. In the case of Image 1, for example, 5 participants stated the image was manipulated based on their belief that match heads cannot be black in colour, which is not the case.

4. 2. Eye Gaze

In the case of Images 10, 11, and 12, participants' eye gaze indicated that increased viewing time as denoted by a greater number of fixations and durations of eye gaze were associated with a greater rate of accuracy in identifying manipulating issues. In addition, although the remaining 5 images did not show this outcome, an overall effect of increased attention associated with increased accuracy was noted when all images were taken as a whole. It is not clear whether this was the result of a person suspecting a manipulated region at the commencement of viewing an image and extending their attention as a result to

seek corroboration with eye gaze or suspecting the manipulation as a result of extended attention of the target area.



(a) (b) (c)

Fig. 3. Comparing 'heatmap' of eye gaze intensity of participants presented with manipulated Image 11.
a) Image 11 as presented,
b) Eye gaze of participants stating: unmanipulated,
c) Eye gaze of participants stating: manipulated.
Key: Red - most viewed, yellow - less viewed, blue - least viewed.

4. 3. Grounded Theory and Eye Gaze

In attempting to understand the reasons that Images 10, 11, and 12 demonstrated a significant difference in attention via eye gaze fixations and durations, we considered the nature of the images. Some common motivations for manipulated images are personal, humour, politics and fraud. It may be that in comparison to more 'mainstream' images that showed a greater similarity in attention (Images 3, 8, 9, 13, 14), Images 10-12 may have been 'harder work' to interpret. For example, although Image 11 may be a political image, it contains an unusual juxtaposition of John Howard and an image of Queen Elizabeth II (Figure 3) that seemed to confound participants, especially those who stated the image was manipulated.

These participants devoted more of their attention and verbal responses to the image of the Queen than to John Howard. This effect on visual attention is illustrated in the heat map of gaze intensity in C of Figure 3, which is greater for the image of the Queen than for John Howard, and also greater in comparison to the same location in image B. The 'unmanipulated' group represented in B of Figure 3 also referred to the image of the Queen 23% less in their comments than the 'manipulated' group shown in C.

It is interesting to note that Image 6 was almost universally misunderstood. Of the 80 participants who viewed this early version of image manipulation from the 19th century, only 3 came close to correctly interpreting that the entire image was a composite derived from 5 negatives; they noted that the man at the window had been added, although this represents only a part of the composition. This may indicate a 'hiding effect' in which a more significant perceived, though incorrect, manipulation (filtering) 'hides' the compositing. This effect is more overt in the case of Image 9; although 49 participants out of 80 stated that the cow has been added to the bonnet of the BMW, only 9 noted in addition that the license plate had been obscured as well, and indeed the blurred license plate received only 2.9% of eye gaze attention by all participants.

5. Conclusion

Based on the outcome of this experiment, we conclude that being presented with an image is in itself insufficient to reliably identify if the image we view has been altered. With an average success rate of 56.0%, we get it wrong 44% of the time. Even when we do successfully identify manipulated images, it is far more difficult for us to tell *what* has been manipulated (27.7%). However, eye gaze tracking shows that for some images and as an overall trend, increased attention (fixation, duration of eye gaze) to manipulated areas of photographs tends to be associated with greater accuracy in deciding if an image is manipulated.

Our eye gaze is a partial reflection of the features of an image that we non-consciously note, but this is not always a predictor of conscious accuracy. Further, more prominent features of an image may obscure our recognition of less obvious manipulations.

In the absence of additional knowledge about the photograph, we resort to the use of logic derived from personal experience which may or may not be relevant, and remain uncertain of our conclusions.

This is problematic because we use images in almost every walk of life from social media to advertising, and from news images to health information, without being able to determine if and how these images have been manipulated.

If viewing a standalone image is not sufficient to allow us to identify if it has been manipulated, even after being pre-familiarised with concepts of image manipulation, then how can we interpret images correctly? It seems plausible that it is necessary for images to be accompanied by a source of additional information such as one or more of an assertion of the status of the image (manipulated or not and how), metadata, context, reference images, or verbal description. We will be investigating this in our further work. If users can be truthfully convinced of the veracity of the images they view online, then this has implications for the design for many forms of human computer interaction via the web.

Acknowledgements

All photographs and photo alterations shown are by Sabrina Caldwell with the following exceptions used under the fair use provisions of copyright law:

- Cow on BMW image used in Image 9. (Web-1).
- Queen Elizabeth II image used in Image 11. (Web-2).
- Brookesia micra chameleon image. Glaser et al, used in Image 1. (Web-3).
- Robinson image "Fading Away" used in Image 6. (Web-4).
- Iranian missile launch image used in Image 8. (Web-5).

References

Caldwell, S., Gedeon, T., Jones, R., & Henschke, M. (2015). Comparing Eye Gaze Tracking to Reported Perceptions of Manipulated and Unmanipulated Digital Images. *Australian Journal of Intelligent Information Processing Systems*, 14(3).

Cromey, D. (2010). Avoiding Twisted Pixels: Ethical Guidelines for the Appropriate Use and Manipulation of Scientific Digital Images. *Science and Engineering Ethics*, 16(4), 639-67.

Ericsson, & Qualcomm. (2013). A Focus on Efficiency. *White Paper*, Facebook

Glaser, B. (2007). Remodeling Grounded Theory. *Historical Social Research*, Supplement. 19, 47-68.

Glaw, F., et al. (2012). Rivaling the World's Smallest Reptiles: Discovery of Miniaturized and Microendemic New Species of Leaf Chameleons (Brookesia). *Northern Madagascar*.

Harding, G., & Bloj, M. (2010). Real and Predicted Influence of Image Manipulations on Eye Movements During Scene Recognition. *Journal of Vision*, 10(2):8, 1-17.

Kress, G. (1996). Internationalisation and Globalisation: Rethinking a Curriculum of Communication. *Comparative Education*, 32(2), Special Number (18), 185-196.

Itti L., & Koch, C. (2000). A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention. *Vision Research*, 40, 1489-1506.

Prasad, H., Wanjari, S., & Rujkumar, P. (2011). Global Manipulation of Digital Images Can Lead to Variation in Cytological Diagnosis. *Journal of Pathology Informatics* 2(20).

Robinson, H. (1860). On Printing Photographic Pictures from Several Negatives. *The British Journal of Photography*, 94.

Treisman, A. & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology* 12, 97-136.

Wheeler, T. (2002). Phototruth or Photofiction?: Ethics and Media Imagery in the Digital Age. *Mahwah, NJ: Lawrence Erlbaum Associates*

Web sites:
Web-1: http://hoaxes.org/weblog/comments/cow_on_hood_of_car, consulted 1 April 2015.
Web-2: http://en.wikipedia.org/wiki/File:Queen_of_canada_wob.jpg, consulted 1 April 2015.
Web-3: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3279364/, consulted 1 April 2015.
Web-4: http://commons.wikimedia.org/wiki/File:Fading_Away.jpg, consulted 1 April 2015.
Web-5: http://www.telegraph.co.uk/news/picturegalleries/worldnews/8982502/Digitally-altered-images-famous-pictures-that-have-been-manipulated-using-Photoshop.html?image=2, consulted 1 April 2015.