InterPARES Trust
Project Report

Title and code:		AA02 User perspectives of trust
Document type:		Final report
Status:		Final
Version:		1.0
Research domain:		Control
Date submitted:		
Last reviewed:		
Author:		InterPARES Trust Project
Writer(s):		Gillian Oliver
Research team:		Peter Bruza, Helen Partridge, Gillian Oliver, Lauren Fell, Morgan Gradwell

Document Control

Version history			
Version	Date	By	Version notes
1.0	2 July 2018	Gillian Oliver	

Table of Contents

Abstract

In an age where any digital image can be manipulated, studying how and why people trust images as well as how likely people are to endorse deceptive images has become a topic of increasing importance. But, how is it human beings decide whether they trust an image? The goal of this study was to identify intersubjective cognitive dimensions which underpin decisions of trust in relation to images, by means of a substantial crowd sourced empirical study.Qualitative data was coded using an axial coding method, with focus placed on themes of the features bearing on decisions of trust.

Findings identified four dimensions: the features of the image, its content, its source, and the participants' own background knowledge. The study also revealed a surprising cognitive effect: In some cases participants confounded the decision of the trustworthiness of an image, with a decision of whether they trust the subject portrayed in the image. It cannot be assumed that digital images will necessarily or automatically be trusted by those viewing these artefacts. There are many socio-cognitive factors in play and a reliable source alone does not consistently determine trustworthiness for users.

User Perspectives of Trust

Research team

Peter Bruza, Queensland University of Technology
Helen Partridge, Queensland University of Technology
Gillian Oliver, Victoria University of Wellington

Graduate Research Assistants: Lauren Fell (QUT) and Morgan Gradwell (VUW).

Background

In archival studies, Donaldson & Conway (2015) identified the following user-centred themes in a qualitative study of a document archive:

- Accuracy: believed to be free of error
- Believability: the extent to which the information appears to be plausible
- Coverage: completeness of the information
- Currency: the degree to which the information is up-to-date
- Objectivity: balance of content
- Stability: the persistence of information, both its presence and contents
- Validity: the use of responsible and accepted practices such as the soundness of the methods used, the inclusion of verifiable data, and the appropriate citation of sources

These, along with Kelton et al.'s (2008, p. 370) similarly identified features of accuracy, objectivity, validity and stability, serve to, as Donaldson and Conway (2015) put it, '*treat the end user, when invoked at all, as the recipient of 'propertied' information, rather than as participants in the formation of trustworthiness'*. This perspective, shared by the present authors, is common in archival studies.

One of the major challenges in properly "invoking" the end user is how to systematically deal with the inherent subjectivity regarding decisions of trust.
Some influences that affect the subjective experiences of the end user can, however, be common across individuals viewing the same image. Greenberg (2013) highlights two cognitive processes that are in operation when a person views an image: those that evaluate the content, resemblance and reference of an image and processes that evaluate the geometric and artistic depiction of an image with regards to reality. Because these processes operate implicitly and pre-consciously, it is possible that they confound in some way if an image is ambiguous. That is, if viewers are challenged with regards to the content, resemblance or reference of an image, it might precipitate a challenge to the geometric or artistic representation of the image and vice versa. It turns out that visual fluency is an important factor in this context.

The concept of visual fluency is based on the principle that any visual stimulus requires cognitive work to process, the more work required, the less fluent the process. Cognitive work includes the evaluation of: content, resemblance, reference of an image; geometric and artistic depictions. Images that cohere with background beliefs on any of these factors are more visually fluent than properties that surprise or confuse us. The amount of cognitive work is reflected in the speed and

accuracy of visual processing as well as in the subjective experience of ease or difficulty of visual judgments (Jacoby, Kelley, & Dywan, 1989; Winkielman, Schwarz, Reber, & Fazendeiro, 2003)

If the visual fluency hypothesis is right, then manipulated images are less detectable the more they conform to largely unconscious rules of visual fluency. Ease of visual processing results in an illusion of accuracy, perhaps because perceptual fluency elicits a feeling of familiarity (Winkielman et al., 2003)—and hence trust.

Given the quite wide range of cognitive processes which influence decisions of trust, it is not surprising that in the few studies that have been conducted, users are not adept at making robust decisions about the trustworthiness of images. One recent study rated users at being "poor to moderate" in their ability to detect manipulated images, and "poor" at identifying what part of the image had been manipulated (Caldwell et al., 2015).

It seems reasonable to assume that when the manipulation does not sufficiently disturb visual fluency, other factors may be employed in the decision making process. One such factor is the source of the image. Studies have shown that the source of information affects users' implicit evaluations (Smith, Houwer & Nosek, 2013, Pornpitakpan, 2004). If the credibility associated with a particular source has the ability to change people's implicit evaluations, this could also be the case with trust. For example, if the source of an image is a reputed institution, the subject might be more willing to trust the image than if it was encountered on social media.

Importantly, trust is also a socio-cognitive phenomenon. The sociological aspect of trusting internet images is particularly acute because frequently images are not associated with reputational source such as Reuters or Getty Images, but appear de-contextualised to agents through anonymising sources such as image search engine results (e.g., Google image search), photo sites (e.g., Flickr), clickbait aggregator sites (e.g., Dose), or personal websites (e.g., Wordpress) usually without attribution. The anonymity of internet images removes relational aspects of trust that define ordinary social interactions (Cook & Gerbasi 2009). On the other hand, the avenues through which anonymous images are found and viewed may have a profound relational aspect – such as if they are shared via social media or through friends. Studies have found, for example, that news is increasingly being preferred to be viewed through social media channels than through professional journalists (Marchi, 2012, Hermida, Fletcher, Korell & Logan, 2012). The relationships users have with their social media circles elicit a personal trust they may not share with a journalist with whom they have never had a personal encounter.

Personality factors can also play an important role in an individual's propensity to trust. It is possible that the user's personality type might sway the decision one way or the other when judging their trust in an image. For example, those with personality traits of extraversion, agreeableness and emotional stability may exhibit a propensity to trust images, as these traits have been identified as correlating with trust (Evans & Revelle, 2008). Conversely, those who are particularly conscientious may be predisposed to be critical and have a propensity to distrust.

In summary, user judgements regarding the trustworthiness of images involve attributes of the image, e.g., the reputation of its source, interaction between the image and human cognitive processes, e.g., determinations of visual fluency and typicality, and sociological factors e.g., relational factors of trust. This is a rather complicated state of affairs. In order to better understand how users transact decisions of trust in relations to images, we conducted a crowdsourced empirical study to identify how these multiple dimensions contribute to a subjective decision of trust.

Goal

The goal of our empirical study was to induce cognitive dimensions that underpin decisions of trust in relation to images. The reason why a cognitive perspective is adopted is that in our view that many of the factors bearing on a decision of trust are ultimately situated in cognition of the user. For example, the source of an image is one of the properties of the image, but is how the user perceives that source which is the pertinent factor. Cognitively this perception is formed at the point of *interaction* with the image.

Hypotheses were:
- A reputed source accompanied with an image will produce higher trustworthiness ratings than one found on a social media source, or with no attributed source;
- People scoring high on scales of "extraversion", "agreeableness" and "emotional stability" personality traits would be more likely to provide higher trustworthiness ratings, whilst those high on "conscientiousness" would provide lower ratings.

Methodology

The methodology that was employed is similar to studies which induced cognitive dimensions underpinning decisions of relevance in relation to documents (Schamber et al., 1990; Barry, 1994; Mizzaro, 1997; Borlund, 2003). The images for the study were carefully selected in order to confound visual fluency to varying degrees.

Participants

Participants consisted of 87 workers using the online crowdsourcing platform, *Amazon Mechanical Turk* (AMT). This is a platform that enables researchers, amongst others, to post experiments and surveys in a form called a HIT (Human Intelligence Task) on a website available to thousands of potential participants to view and complete. Participants involved in the present study were able to view any HIT before agreeing to participate, and were paid a small amount (such as 50 – 60c) per HIT. AMT allows workers of a certain skill, ability or reputation to be specified for a HIT. In this experiment, workers of at least 95% or greater approval rating were selected to balance worker quality with the need to attract a sufficient number of participants. No demographic data was taken of participants, however, the typical demographics of AMT workers are generally known. Around 50% are from the United States, 40% from India, and 10% from other countries (Paolacci et. al., 2010). Workers using this platform are predominantly female if residing in the United States, and predominantly male if residing in India (Paolacci et. al., 2010). Because the experiments were presented in English, it was assumed that workers choosing to participate would be proficient in this language. Data collected from participants who did not respond to written segments of the experiment in English, or who were deemed not to have understood instructions based on the relevancy/quality of their written responses, were excluded from the study. In total, 3 workers were excluded on this basis.

Materials

The *Ten Item Personality Inventory (TIPI)* developed by Gosling Rentfrow & Swann (2003) was designed as a short measure of the five-factor model of personality. The five-factor model stipulates that personality can be broken into five scales; openness to experiences, conscientiousness, extraversion, agreeableness and emotional stability. Participants were asked to what extent each item reflects how they see themselves on a 7-point scale. Each personality trait had both a positive and reverse-scored item, for example, the two items measuring extraversion

were *extraverted, enthusiastic*, and *reserved, quiet* (the latter being reverse-scored). Gosling, Rentfrow & Swann (2003) found the measure to have strong test-retest reliability ($r = .72$) as well as convergent validity (mean $r = .77$).

Images were obtained using a *Google images* search, and were freely available for use. All images used were chosen because they represented an unusual or unexpected depiction of the subject they portrayed. The aim of image selection was to present images that challenged participants' implicit representations of the subjects, creating visual disfluency and triggering a predisposition to distrust. The group also included a mix of digitally altered and unaltered images, as well as a mix of sources displayed underneath the image. The source condition consisted of 3 levels: a source of social media origin (Facebook), a reputed source (either the Museum of Natural History in South Africa, or the Museum of Modern History in South Africa, depending on the nature of the image), and a lack of a source. Using a mixture of geometrically 'real' and 'fake' images, mix of source reputation, along with visually disfluent subjects, the aim was to assess how these influencers of trust – representative, reputational and geometric – interact in cognitive decision making around trust.

The following images were used: a photograph of Russian President Vladmir Putin (unaltered), a photograph of a frill shark (unaltered), a picture of a man running from an explosion (digitally altered), a photograph of a mountain with an image of deep space taken by the Hubble Telescope superimposed in the background (digitally altered), and a photograph of a train derailment at Montparnasse Station in 1895 (unaltered).

Procedure

Participants were instructed to peruse each HIT before agreeing to participate. Each HIT began with the 10-item personality questionnaire, followed by a definition of trustworthiness and a set of questions. The definition of trustworthiness given to participants was as follows:

Trustworthiness can be defined as an accurate representation of a situation, person or object

This was followed by 5 questions, each of which began with an image. Each image was contained within a red border and presented in the same order in each HIT. The source condition was presented between-subjects, where all 5 images a participant viewed contained the same type of source underneath. Following each image was a question, which read:
Taking into account the image itself, please indicate the level to which you judge the trustworthiness *of the above image based on features* inside the red box.

This was followed by a Likert scale response system ranging from *very untrustworthy* (1) to *very trustworthy* (5) as well as a textbox asking participants to explain the reasons behind their decisions. After completion of the experiment, participants were given the opportunity to choose to submit their HITs.

Analysis

After all data had been gathered, qualitative data was coded using an axial coding method, with focus placed on themes of the features bearing on decisions of trust. Coding revealed four themes:
1. Features of the image itself (e.g. "*The person in the foreground does not seem to blend with the image in the background*")
2. Content/Subject of the image (e.g. "*All wild animals are untrustworthy. Also look at his teeth. He acts on instinct alone.*")

3. Source below the image (e.g. *"it is from a museum, so it seems to be trustworthy"*)
4. Prior knowledge (e.g. "*Could be faked, but the auroras are spectacular and always look fake."*)

Findings

This article set out to explore the cognitive decision space for deciding the trustworthiness of images. The qualitative analysis revealed the following four themes, which we put forward as corresponding to underlying dimensions of this decision space:

- Features of the image itself (e.g., '*The person in the foreground does not seem to blend with the image in the background'*)
- Content/Subject of the image (e.g., '*All wild animals are untrustworthy. Also look at his teeth. He acts on instinct alone.'*)
- Image source (e.g., *'it is from a museum, so it seems to be trustworthy'*)
- Prior knowledge (e.g., *'Could be faked, but the auroras are spectacular and always look fake.'*)

The dimensions of trustworthiness allow a comparison of user-centred themes of trustworthiness induced from a qualitative study of archival documents (Donaldson & Conway, 2015). Recall that the following dimensions were induced from their study:

- Accuracy: believed to be free of error
- Believability: the extent to which the information appears to be plausible
- Coverage: completeness of the information
- Currency: the degree to which the information is up-to-date
- Objectivity: balance of content
- Stability: the persistence of information, both its presence and contents
- Validity: the use of responsible and accepted practices such as the soundness of the methods used, the inclusion of verifiable data, and the appropriate citation of sources

Donaldson & Conway also uncovered the following emergent themes:

- Perceived authenticity: Is it fake?
- Inaccurate information: conceptualizing documents as being trustworthy despite containing inaccurate information
- Primary or first hand evidence: the extent to which the document is primary or first-hand
- Document legibility or readability
- Document's perceived proper form.

Although the preceding themes "coverage", "readability", "proper form" and "validity" relate to the information being in the form of a document, it is nevertheless possible to draw some comparisons with the four dimensions that were induced from the present experiment based on information in the form of images.

Donaldson & Conway's theme of "authenticity" relates to the "image features" dimension as the latter comprises the identification of areas of the image that look fake or suspicious. A contrast can also be drawn with the theme of "accuracy", but in the case of images it is not freedom from error that underpins the decision but issues such as whether the image is deemed an "accurate enough" portrayal of the subject of the image.

In both documents and images "believability" hinges on determining the plausibility of the content. The present experiment uncovered that in the case of images prior knowledge is an important component of that determination.

An important and unexpected finding is that the subject of an image has a considerable effect on the judgement of trustworthiness of that image, particularly if the viewer distrusts the subject portrayed by the image. In some ways this finding is the converse of an aspect of the "validity" uncovered by Donaldson & Conway where a document would be regarded as trustworthy even if its content is inaccurate, or incorrect (i.e., untrustworthy). In the present experiment, an image could be deemed *untrustworthy* simply because the subject (i.e., the content) of the image is deemed untrustworthy. If this finding is reliable, it has implications for using people to judge the credibility of images, as it appears that for some images the decision about the trustworthiness of the image is being confounded with a decision regarding the trustworthiness of the *subject* of the image.

More broadly the findings of this study suggest that it cannot be assumed that digital images will necessarily or automatically be trusted by those viewing these artefacts. There are many socio-cognitive factors in play and a reliable source alone does not consistently determine trustworthiness for users. Library, archival and museum advocacy and outreach programmes concentrating on increasing the visibility of collections by curating online exhibitions need to develop awareness of the nuanced and complex factors influencing user decision making about trust issues. This is a particularly significant issue for the archival professional community given the critical nature of ensuring the trustworthiness of records (see, for example, Duranti and Rogers, 2012). The findings suggest that it is overly simplistic to assume that by establishing a link with a cultural heritage institution trust will automatically ensue.

Conclusions

Findings indicate the need for much more research in this area, comparing for instance the impact of making collections available on different platforms. In our view, users are not recipients of 'propertied' information regarding the object, but are participants in the formation of those properties while interacting with that object. Consequently, a better understanding is required of how human subjectivity is involved when these properties are formed at the point of interaction. In short, we advocate that the cultural heritage sector should have a more substantive dialogue with the field of cognitive science.

References

Barry, C. (1994). User-defined relevance criteria: an exploratory study. *J. Am. Soc. Inform. Sci.* 45, 145–159.

Borlund, P. (2003). The concept of relevance in IR. *J. Am. Soc. Inform. Sci. Tech.* 54, 913–925. doi: 10.1002/asi.10286

Bruza, P.D., Busemeyer, J. & Wang, Z. (2015) Quantum cognition: a new theoretical approach to psychology. *Trends in Cognitive Sciences*, 19(7), 383-393.

Bruza, P.D., & Hoenkamp, E.C.M. (2017) Thinking fast and slow in autonomous systems: a quantum cognitive approach to reinforce trust between human and machine In Abbas, H.A., Scholz, J.B., & Reid, D.J. (eds) Foundations of Trusted Autonomy. Springer.

Caldwell, S., Gedeon, T., Jones, R. and Copeland, L. (2015). "Imperfect Understandings: A Grounded Theory and Eye Gaze Investigation of Human Perceptions of Manipulated and Unmanipulated

Digital Images" in *Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (ECSS 2015)*

Carlson, M. (2009). The Reality of a Fake Image; News norms, photojournalistic craft, and Brian Walski's fabricated photograph. *Journalism Practice*, *3*(2), 125–139.

Cook, K. S., & Gerbasi, A. (2009). Chapter 10: Trust. In P. Hedström & P. S. Bearman (Eds.), *Oxford Handbook of Analytical Sociology* (pp. 218-241): Oxford University Press.

Donaldson, D.R & Conway, P. (2015). User conceptions of trustworthiness for digital archival documents. *Journal of the Association for Information Science and Technology*, 66(12), 2427-2444.

Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *The Journal of Marketing*, 35-51.

Duranti, L., & Rogers, C. (2012). Trust in digital records: An increasingly cloudy legal area. *Computer Law & Security Review*, *28*(5), 522–531.

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*(6), 1585-1593.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504-528.

Greenberg, G. (2013). Beyond resemblance. *Philosophical Review* 122(2): 215-287

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend: Decoding the social media news consumer. *Journalism Studies*, *13*(5-6), 815-824.

InterPARES. (2008). InterPARES 3 Project: International Research on Permanent Authentic Records in Electronic Systems. Retrieved July 22, 2014, from http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=t&term=55

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Kelton, K., Fleischmann, K.R., & Wallace, W.A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, 59(3), 363–374.

Los Angeles Times. (2005). Los Angeles Times Ethics Guidelines. Retrieved from http://media.trb.com/media/acrobat/2005-07/18479691.pdf

Lum, J. (2010, July 5). Controversy Crops Up Over Economist Cover Photo. PetaPixel. Retrieved from http://petapixel.com/2010/07/05/controversy-crops-up-over-economist-cover-photo/

Lum, J. (2010, July 19). Getty Photographer Terminated Over Altered Golf Photo. PetaPixel. Retrieved from http://petapixel.com/2010/07/19/getty-photographer-terminated-over-altered-golf-photo/

Marchi, R. (2012). With Facebook, blogs, and fake news, teens reject journalistic "objectivity". *Journal of Communication Inquiry*, *36*(3), 246-262.

Mizzaro, S. (1997). Relevance: the whole history. *JASIS.* 48(9), 810–832.

National Press Photography Association. (2012). NPPA Code of Ethics. Retrieved May 19, 2014, from https://nppa.org/code_of_ethics

Paolacci, G, Chandler, J. & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgement and Decision Making*, 5(5), 411-419.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2), 243-281.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*(3), 338-342.

Schamber, L., Eisenberg, M., and Nilan, M. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing & Management.* 26(6), 755–775.

Shane, S. and Goel, V. (2017). *Fake Russian Facebook Accounts Bought $100,000 in Political Ads* https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html

Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the Source Persuasion of Implicit Evaluations Is Moderated by Source Credibility. *Personality and Social Psychology Bulletin*, *39*(2), 193-205.

Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, *1124*(1), 208-224.

Winkielman, P., Schwarz, N., Reber, R., & Fazendeiro, T. A. (2003). Affective and Cognitive Consequences of Visual Fluency: When Seeing is Easy on the Mind. In R. Baatra & L. Scott

(Eds.), *Persuasive imagery: A consumer response perspective* (pp. 75-89). Mahwah, NJ: Lawrence Erlbaum.