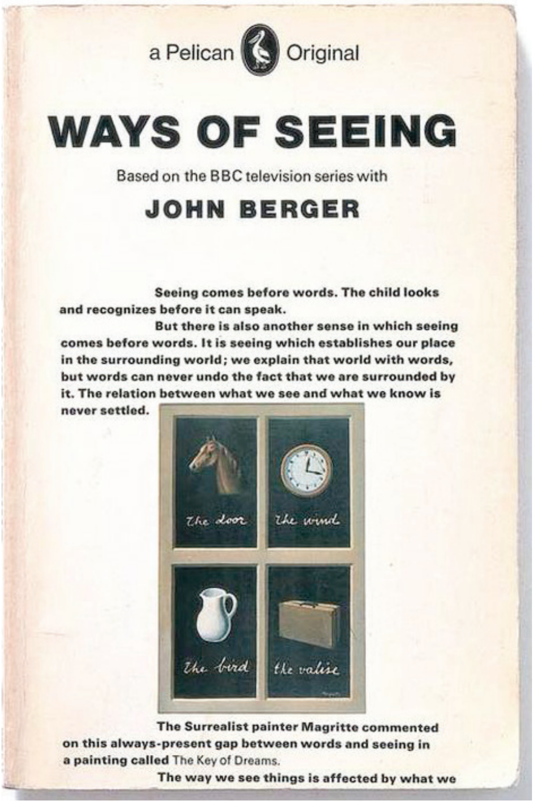
Geoff Cox

WAYS OF MACHINE SEEING: AN INTRODUCTION

APRJA Volume 6, Issue 1, 2017
ISSN 2245-7755



a Pelican Original

WAYS OF SEEING

Based on the BBC television series with

JOHN BERGER

Seeing comes before words. The child looks and recognizes before it can speak.

But there is also another sense in which seeing comes before words. It is seeing which establishes our place in the surrounding world; we explain that world with words, but words can never undo the fact that we are surrounded by it. The relation between what we see and what we know is never settled.

The Surrealist painter Magritte commented on this always-present gap between words and seeing in a painting called The Key of Dreams.

The way we see things is affected by what we

Figure 1: The Cover of *Ways of Seeing* by John Berger (1972). Image from Penguin Books.

You are looking at the front cover of the book *Ways of Seeing* written by John Berger in 1972.[1] The text is the script of the TV series, and if you've seen the programmes, you can almost hear the distinctive pedagogic tone of Berger's voice as you read his words: "The relation between what we see and what we know is never settled."[2]

The image by Magritte on the cover further emphasises the point about the deep ambiguity of images and the always-present difficulty of legibility between words and seeing.[3] In addition to the explicit reference to the "artwork" essay by Walter Benjamin,[4] the TV programme employed Brechtian techniques, such as revealing the technical apparatus of the studio; to encourage viewers not to simply watch (or read) in an easy way but rather to be forced into an analysis of elements of "separation" that would lead to a "return from alienation".[5] Berger further reminded the viewer of the specifics of the technical reproduction in use and its ideological force in a similar manner:

> *But remember that I am controlling and using for my own purposes the means of reproduction needed for these programmes […] with this programme as with all programmes, you receive images and meanings which are arranged. I hope you will consider what I arrange but please remain skeptical of it.*

That you are not really looking at the book as such but a scanned image of a book — viewable by means of an embedded link to a server where the image is stored — testifies to the ways in which what, and how, we see and know is further unsettled through complex assemblages of elements. The increasing use of relational machines such as search engines is a good example of the ways in which knowledge is filtered at the expense of the more specific detail on how it was produced. Knowledge is now produced in relation to planetary computational infrastructures in which other agents such as algorithms generalise massive amounts of (big) data.[6]

Clearly algorithms do not act alone or with magical (totalising) power but exist as part of larger infrastructures and ideologies. Some well-publicised recent cases have come to public attention that exemplify a contemporary politics (and crisis) of representation in this way, such as the Google search results for "three black teenagers" and "three white teenagers" (mug shots and happy teens at play, respectively).[7] The problem is one of learning in its widest sense, and "machine learning" techniques are employed on data to produce forms of knowledge that are inextricably bound to hegemonic systems of power and prejudice.

There is a sense in which the world begins to be reproduced through computational models and algorithmic logic, changing what and how we see, think and even behave. Subjects are produced in relation to what algorithms understand about our intentions, gestures, behaviours, opinions, or desires, through aggregating massive amounts of data (data mining) and machine learning (the predictive practices of data mining).[8] That machines learn is accounted for through a combination of calculative practices that help to approximate what will likely happen through the use of different algorithms and models. The difficulty lies in to what extent these generalisations are accurate, or to what degree the predictive model is valid, or "able to generalise" sufficiently well. Hence the "learners" (machine learning algorithms), although working at the level of generalisation, are also highly contextual and specific to the fields in which they operate in a coming together of what Adrian Mackenzie calls a "play of truth and falsehood".[9]

Thus what constitutes knowledge can be seen to be controlled and arranged in new ways that invoke Berger's earlier call for skepticism. Antoinette Rouvroy is similarly concerned that algorithms begin to define what counts for knowledge as a further case of subjectivation, as we are unable to substantively intervene in these processes of how knowledge is produced.[10] Her claim is that knowledge is delivered "without truth" through the increasing use of machines that filter it through the use of search engines that have no interest in content as such or detail on how knowledge is generated. Instead they privilege real-time relational infrastructures that subsume the knowledge of workers and machines into generalised assemblages as techniques of "algorithmic governmentality".[11]

In this sense, the knowledge produced is bound together with systems of power that are more and more visual and hence ambiguous in character. And clearly computers further complicate the field of visuality, and ways of seeing, especially in relation to the interplay of knowledge and power. Aside from the totalizing aspects (that I have outlined thus far), there are also significant "points of slippage or instability" of epistemic authority,[12] or what Berger would have no doubt identified as the further unsettling of the relations between seeing and knowing. So, if algorithms can be understood as seeing, in what sense, and under what conditions? Algorithms are ideological only inasmuch as they are part of larger infrastructures and assemblages.

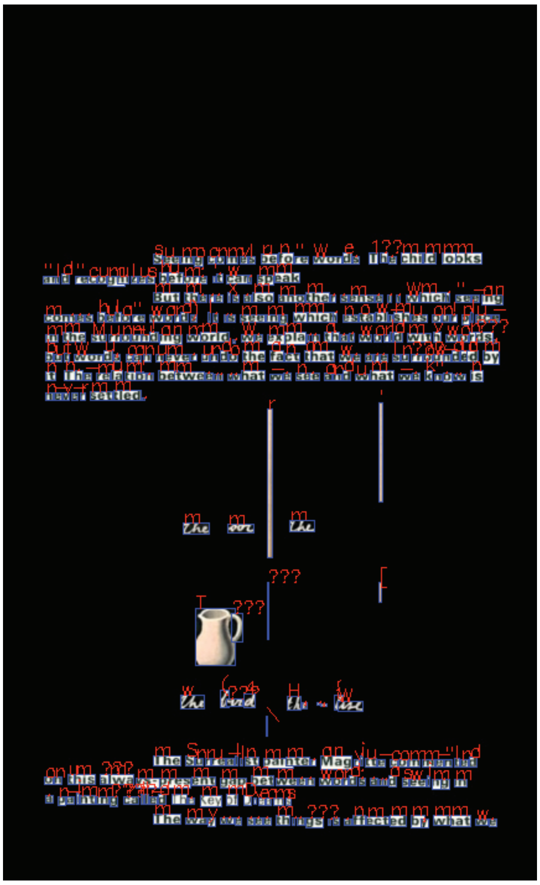
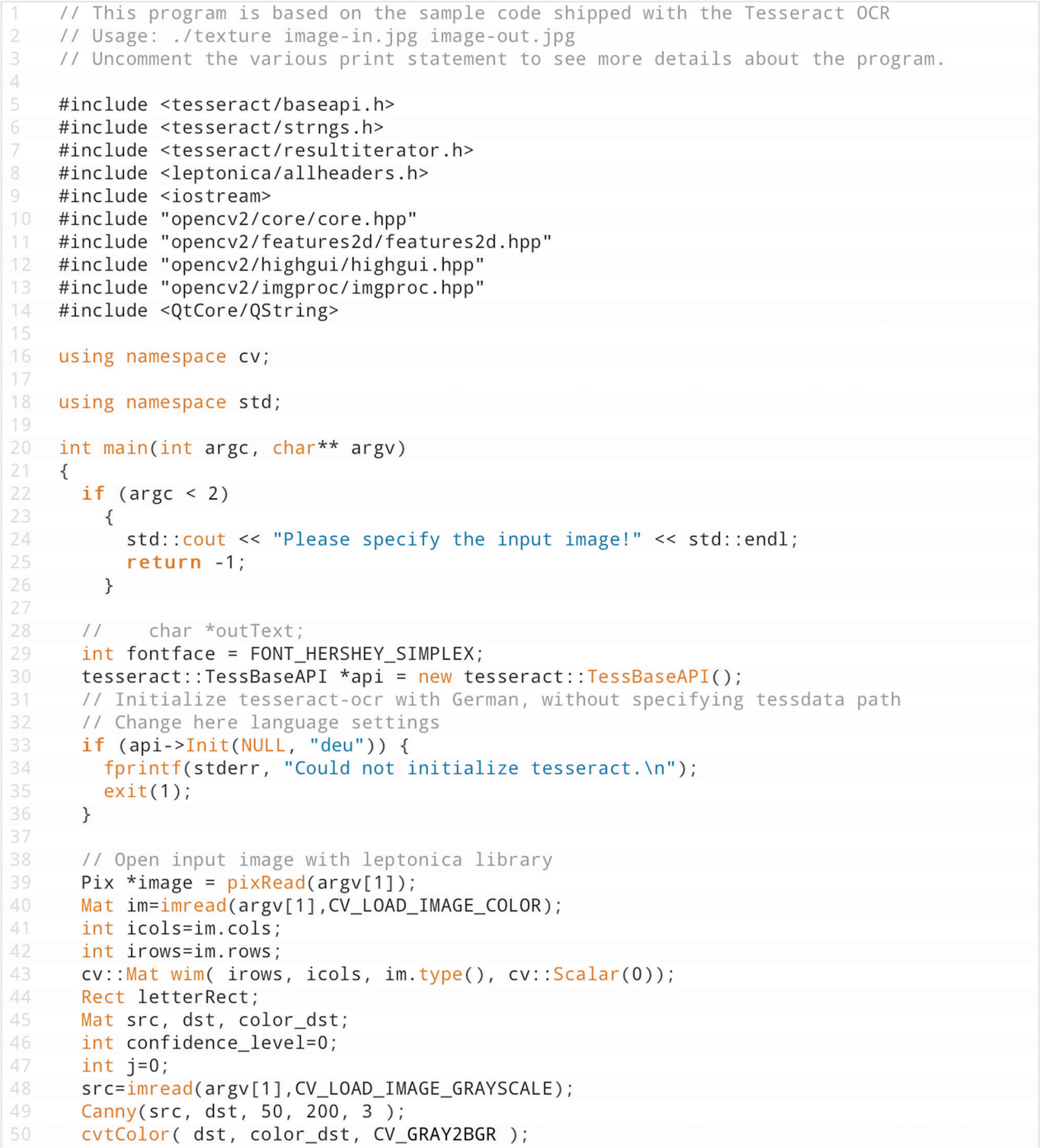

Seeing comes before words. The child looks
and recognizes before it can speak.
But there is also another sense in which seeing
comes before words. It is seeing which establishes our place
in the surrounding world; we explain that world with words,
but words can never undo the fact that we are surrounded by
it. The relation between what we see and what we know is
never settled.
The Surrealist painter Magritte commented
on this always-present gap between words and seeing in
a painting called The Key of Dreams
The way we see things is affected by what we

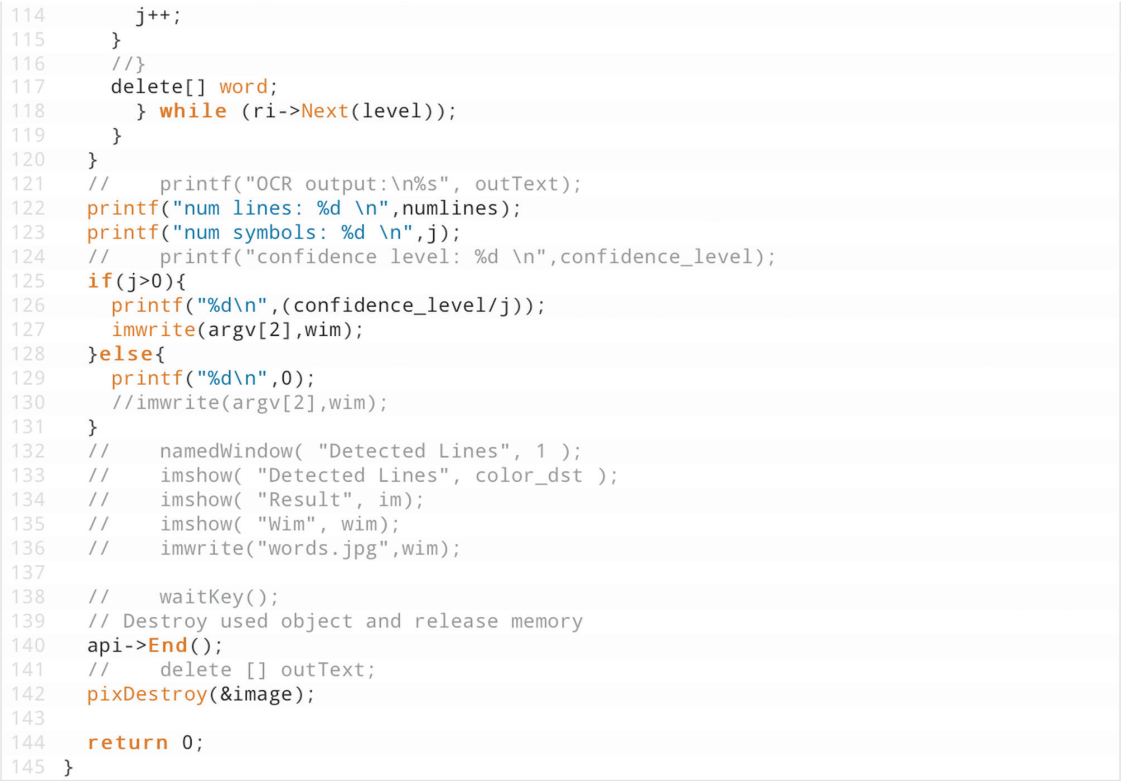
Figure 2: The *Ways of Seeing* book cover image seen through an optical character recognition program. Created by SICV.

But to ask whether machines can see or not is the wrong question to ask, rather we should discuss how machines have changed the nature of seeing and hence our knowledge of the world.[13] In this we should not try to oppose machine and human seeing but take them to be more thoroughly entangled — a more “posthuman” or “new materialist” position that challenges the onto-epistemological character of seeing — and produces new kinds of knowledge-power that both challenges as well as extends the anthropomorphism of vision and its attachment to dominant forms of rationality. Clearly there are other (nonhuman) perspectives that also illuminate our understanding of the world. This pedagogic (and political) impulse is perfectly in keeping with Ways of Seeing and its project of visual literacy.[14] What is required is an expansion of this ethic to algorithmic literacy to examine how machine vision unsettles the relations between what we see and what we know in new ways.



```
// This program is based on the sample code shipped with the Tesseract OCR
// Usage: ./texture image-in.jpg image-out.jpg
// Uncomment the various print statement to see more details about the program.

#include <tesseract/baseapi.h>
#include <tesseract/strngs.h>
#include <tesseract/resultiterator.h>
#include <leptonica/allheaders.h>
#include <iostream>
#include "opencv2/core/core.hpp"
#include "opencv2/features2d/features2d.hpp"
#include "opencv2/highgui/highgui.hpp"
#include "opencv2/imgproc/imgproc.hpp"
#include <QtCore/QString>

using namespace cv;

using namespace std;

int main(int argc, char** argv)
{
  if (argc < 2)
    {
      std::cout << "Please specify the input image!" << std::endl;
      return -1;
    }

  //    char *outText;
  int fontface = FONT_HERSHEY_SIMPLEX;
  tesseract::TessBaseAPI *api = new tesseract::TessBaseAPI();
  // Initialize tesseract-ocr with German, without specifying tessdata path
  // Change here language settings
  if (api->Init(NULL, "deu")) {
    fprintf(stderr, "Could not initialize tesseract.\n");
    exit(1);
  }

  // Open input image with leptonica library
  Pix *image = pixRead(argv[1]);
  Mat im=imread(argv[1],CV_LOAD_IMAGE_COLOR);
  int icols=im.cols;
  int irows=im.rows;
  cv::Mat wim( irows, icols, im.type(), cv::Scalar(0));
  Rect letterRect;
  Mat src, dst, color_dst;
  int confidence_level=0;
  int j=0;
  src=imread(argv[1],CV_LOAD_IMAGE_GRAYSCALE);
  Canny(src, dst, 50, 200, 3 );
  cvtColor( dst, color_dst, CV_GRAY2BGR );
```

```
 vector<Vec4i> lines;
 HoughLinesP( dst, lines, 1, CV_PI/180, 80, 80, 10 );
 int numlines=0;
 for( size_t i = 0; i < lines.size(); i++ )
  {
   double Angle = atan2(lines[i][3]- lines[i][1], lines[i][2]- lines[i][0]) * 180.0 / CV_PI;
   if(Angle>=0 && Angle <=5){
  numlines++;
  line( wim, Point(lines[i][0], lines[i][1]),
     Point(lines[i][2], lines[i][3]), Scalar(0,0,255), 1, 8 );
   }
  }
 if(numlines>0){
  //Mat screen=cvCreateMat(im.rows,im.cols,im.type());
  api->SetImage(image);
  api->Recognize(0);
  //outText = api->GetUTF8Text();
  tesseract::ResultIterator* ri = api->GetIterator();
  tesseract::PageIteratorLevel level = tesseract::RIL_SYMBOL;

  if (ri != 0) {
   do {
  const char* word = ri->GetUTF8Text(level);
  float conf = ri->Confidence(level);
  int x1, y1, x2, y2;
  ri->BoundingBox(level, &x1, &y1, &x2, &y2);
  //       printf("word: '%s';  \tconf: %.2f; BoundingBox: %d,%d,%d,%d;\n",
  //          word, conf, x1, y1, x2, y2);
  bool ignore;
  int psize,fid;
  ri->WordFontAttributes(&ignore,&ignore,&ignore,&ignore,&ignore,&ignore,&psize,&fid);
  //     printf("point size: %d font id: %d",psize, fid);
  if(conf>50){
   QString utfw = QString::fromUtf8(word);
   //         if(utfw.size()>4){
   //      //     printf("word: '%s'  \tlength: %d\n",word,utfw.length());
   letterRect=CvRect();
   letterRect.x=x1;
   letterRect.y=y1;
   letterRect.width=(x2-x1);
   letterRect.height=(y2-y1);
   Mat letterim;
   try{
    letterim=im(letterRect).clone();
    //             im.copyTo(wim(Rect(letterRect.x, letterRect.y, letterRect.width, letterRect.height)));

    cv::Rect roi( cv::Point( letterRect.x, letterRect.y), cv::Size( letterRect.width, letterRect.height));
    //             printf("x %d y %d width %d height %d \n",letterRect.x, letterRect.y,letterRect.width, letterRect.height);
    //             printf("cols %d rows %d\n",wim.cols,wim.rows);
    cv::Mat destinationROI = wim( roi );
    letterim.copyTo( destinationROI );
    QString qs = QString("words/word_%1_").arg(word);
    qs.append(QString("%1.jpg").arg(j));
    imwrite(qs.toStdString(),letterim);
   }
   catch( cv::Exception& e )
    {
     const char* err_msg = e.what();
     //             std::cout << "exception caught: " << err_msg << std::endl;
    }
   rectangle(wim,Point(x1, y1),Point(x2,y2),Scalar(255,0,0),1,8);
   putText(wim,word,Point(x1,y1),fontface,0.5,CV_RGB(255,0,0), 1, 8);
   confidence_level+=conf;
```



```
      j++;
    }
    //}
    delete[] word;
      } while (ri->Next(level));
    }
  }
  //    printf("OCR output:\n%s", outText);
  printf("num lines: %d \n",numlines);
  printf("num symbols: %d \n",j);
  //    printf("confidence level: %d \n",confidence_level);
  if(j>0){
    printf("%d\n",(confidence_level/j));
    imwrite(argv[2],wim);
  }else{
    printf("%d\n",0);
    //imwrite(argv[2],wim);
  }
  //    namedWindow( "Detected Lines", 1 );
  //    imshow( "Detected Lines", color_dst );
  //    imshow( "Result", im);
  //    imshow( "Wim", wim);
  //    imwrite("words.jpg",wim);

  //    waitKey();
  // Destroy used object and release memory
  api->End();
  //    delete [] outText;
  pixDestroy(&image);

  return 0;
}
```

Figure 3: Code by The Scandinavian Institute for Computational Vandalism.

Notes

[1] This essay was first commissioned by The Photographers Gallery for their *Unthinking Photography* series, https://unthinking.photography/themes/machine-vision/ways-of-machine-seeing. The title is taken from a workshop organised by the Cambridge Digital Humanities Network, convened by Anne Alexander, Alan Blackwell, Geoff Cox and Leo Impett, and held at Darwin College, University of Cambridge, 11 July 2016, http://www.digitalhumanities.cam.ac.uk/Methods/waysofmachineseeing; a subsequent workshop, *Ways of Machine Seeing 2017*, is a two-day workshop organised by the Cambridge Digital Humanities Network, and CoDE (Cultures of the Digital Economy Research Institute) and Cambridge Big Data, to be held 26-28 June 2017, http://www.digitalhumanities.cam.ac.uk/Methods/woms2017/woms2017CFP.

[2] *Ways of Seeing*, Episode 1 (1972), https://www.youtube.com/watch?v=0pDE4VX_9Kk. The 1972 BBC four-part television series of 30-minute films was created by writer John Berger and producer Mike Dibb. Berger's scripts were adapted into a book of the same name, published by Penguin also in 1972. The book consists of seven numbered essays: four using words and images; and three essays using only images. See https://en.wikipedia.org/wiki/Ways_of_Seeing.

[3] René Magritte, *The Key of Dreams* (1930), https://courses.washington.edu/hypertxt/cgi-bin/book/wordsinimages/key-dreams.jpg. Aside from the work of Magritte, Joseph Kosuth's *One and Three Chairs*

(1965) comes to mind, that makes a similar point in presenting a chair, a photograph of the chair, and an enlarged dictionary definition of the word "chair", https://en.wikipedia.org/wiki/One_and_Three_Chairs.

[4] The first section of the programme/book is acknowledged to be largely based on Benjamin's essay "The Work of Art in the Age of Mechanical Reproduction" (1936), https://www.marxists.org/reference/subject/philosophy/works/ge/benjamin.htm.

[5] The idea is that "separation" produces a disunity that is disturbing to the viewer/reader — Brecht's "alienation-effect" (Verfremdungeffekt) — and that this leads to a potential "return from alienation". See https://en.wikipedia.org/wiki/Distancing_effect.

[6] To give a sense of scale and its consequences, Facebook has developed the face-recognition software DeepFace. With over 1.5 billion users that have uploaded more than 250 billion photographs, it is allegedly capable of identifying any person depicted in a given image with 97% accuracy. See https://research.facebook.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/.

[7] Antoine Allen "The 'three black teenagers' search shows it is society, not Google, that is racist", *The Guardian* (10 June 2016), https://www.theguardian.com/commentisfree/2016/jun/10/three-black-teenagers-google-racist-tweet.

[8] Adrian Mackenzie, "The Production of Prediction: What Does Machine Learning Want?," *European Journal of Cultural Studies*, 18, 4–5 (2015): 431.
[9] Mackenzie, "The Production of Prediction," 441.

[10] See, for instance, Antoinette Rouvroy's "Technology, Virtuality and Utopia: Governmentality in an Age of Autonomic Computing," in *The Philosophy of Law Meets the Philosophy of Technology: Computing and Transformations of Human Agency*, eds. Mireille Hildebrandt and Antoinette Rouvroy (London: Routledge, 2011), 136–157.

[11] This line of argument is also close to what Tiziana Terranova has called an "infrastructure of autonomization", making reference to Marx's views on automation, particularly in his "Fragment on Machines", as a description of how machines subsume the knowledge and skill of workers into wider assemblages. Tiziana Terranova, "Red Stack Attack! Algorithms, capital and the automation of the common", *Effimera* (2014), accessed August 24, 2016, http://effimera.org/red-stack-attack-algorithms-capital-and-the-automation-of-the-common-di-tiziana-terranova/.

[12] Mackenzie, "The Production of Prediction," 441.

[13] I take this assertion from Benjamin once more, who considered the question of whether film or photography to be art secondary to the question of how art itself has been radically transformed:
"Earlier much futile thought had been devoted to the question of whether photography is an art. The primary question — whether the very invention of photography had not transformed the nature of art — was not raised. Soon the film theoreticians asked the same ill-considered question with regard to film." https://www.marxists.org/reference/subject/philosophy/works/ge/benjamin.htm.

[14] Berger was associated with The Writers and Readers Publishing Cooperative, aiming to "advance the needs of cultural literacy, rather than cater to an 'advanced' [academic] but limited readership" (From the Firm's declaration of intent). In this sense it draws upon the Marxist cultural materialism of Raymond Williams and Richard Hoggart's *The Uses of Literacy* (1966).

Works cited

Allen, Antoine. "The 'three black teenagers' search shows it is society, not Google, that is racist." *The Guardian* (10 June 2016), https://www.theguardian.com/commentisfree/2016/jun/10/three-black-teenagers-google-racist-tweet. Web.

Benjamin, Walter. "The Work of Art in the Age of Mechanical Reproduction." (1936). https://www.marxists.org/reference/subject/philosophy/works/ge/benjamin.htm. Print.

Berger, John. *Ways of Seeing*. London: Penguin, 1972. Print.

Cox, Geoff. "Ways of Machine Seeing." *Unthinking Photography*. London: The Photographers Gallery, 2016. https://unthinking.photography/themes/machine-vision/ways-of-machine-seeing. Web.

Mackenzie, Adrian. "The Production of Prediction: What Does Machine Learning Want?" *European Journal of Cultural Studies*, 18, 4–5 (2015): 431. Print.

Rouvroy, Antoinette. "Technology, Virtuality and Utopia: Governmentality in an Age of Autonomic Computing." Eds. Mireille Hildebrandt and Antoinette Rouvroy. *The Philosophy of Law Meets the Philosophy of Technology: Computing and Transformations of Human Agency*. London: Routledge, 2011. 136–157. Print.

Terranova, Tiziana. "Red Stack Attack! Algorithms, capital and the automation of the common." *Effimera* (2014). http://effimera.org/red-stack-attack-algorithms-capital-and-the-automation-of-the-common-di-tiziana-terranova/. Accessed August 24, 2016. Web.