Horizon Scanning Series

The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing

AI in AU

This input paper was prepared by Professor Robert C Williamson (Australian National University)

Suggested Citation

Williamson, R C (2018). AI in AU. Input paper for the Horizon Scanning Project "The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing" on behalf of the Australian Council of Learned Academies, www.acola.org.

The views and opinions expressed in this report are those of the author and do not necessarily reflect the opinions of ACOLA.

AI ∈ AU

Input offered to the ACOLA working group on Artificial Intelligence

Robert C. Williamson
Australian National University
23 July 2018

I was asked to answer, with references, the following questions in respect of AI (Artificial Intelligence) in the context of a report to the Australian Commonwealth Science Council.

1. What advances in the technology and capabilities of AI in ML can we expect in next decade?
2. In the next decade, what new applications of these technologies will we start to see?
3. What is needed in research infrastructure, legislation, etc. to enable development and adoption of these technologies?

I will answer these questions, but discursively. I think the detour is important, critically because I think you are asking the wrong questions, as I attempt to argue below. *This is perhaps the most significant thing that I can offer: do not ask questions 1 or 2 – you will get answers, but they will be wrong!*

The first two are questions of technological prediction. The third is a much broader question concerning the crafting of an environment that facilitates the development of the technologies. I answer the questions in these two groups. At the end of the document I offer some additional input on the basis of the draft table of contents for the ACOLA report that is being constructed. But to start with I offer some thoughts on how to delineate what AI is, what it is isn't, and what it might be. Much of my thinking draws upon the 2015 ACOLA report *Technology and Australia's Future*[1] which I know inside-out, and which I commend to the working group, and cite from extensively below. It is the basis for my claim that your first two questions, and parts of the draft report structure, is misguided. It is somewhat disappointing that some of the key conclusions, supported by overwhelming evidence, in a report published by ACOLA less than three years ago seem to be entirely ignored in the present exercise!

What is AI … for the purposes at hand

AI is a name for a loosely defined collection of technologies, or collective technology. Technologies can be (and need to be) viewed at different levels of abstraction – there is (for example) the

[1] Robert C. Williamson, Michelle Nic Raghnaill, Kirsty Douglas and Dana Sanchez, *Technology and Australia's future: New technologies and their role in Australia's security, cultural, democratic, social and economic systems,* Australian Council of Learned Academies, September 2015. Available online at http://users.cecs.anu.edu.au/~williams/TAAF.pdf Working papers (which provide substantial additional detail) are available at https://acola.org.au/wp/new-technologies-contributing-reports/ The present notes are written in a bit of a rush, without the advantage of colleagues to review and refine the text. The ACOLA report, in contrast, was well polished.

technology of continuous welded rail versus the technology of railways[2]. The AI collective is as broad as railways, electricity, the automobile, the computer and radio/television. These are usefully described as "General Purpose Technologies" and are widely held to be the major cause of economic growth in the past two centuries[3] although the effects are often long-delayed[4].

Claiming AI is a technology immediately begs the question of what is a technology? The answer is commonly taken to be identified with technological artefacts – the things one can hold in one's hand, or at least touch. But it is far more valuable (and more accurate) I believe to view technology as useful knowledge[5]. This distinction matters significantly when one comes to the interventions one might wish to make; see the final section of this document.

A common cause for confusion in thinking about the impact of technologies is a confounding of the means and the use, or the problem to be solved. The distinction was drawn long ago by sociologist Seabury Gillfillan in section 3.4 of a report to the US President from 1937[6]. Thus, rather than thinking of AI in terms of the techniques used (e.g. "deep" learning), it is likely to be more helpful to focus on the technical problems that are being solved, or attempted to be solved. These are actually easy to describe, being primarily decision and reasoning problems, which includes statistical pattern recognition, planning and optimization, object detection, causal inference. All such problems can (but often are not) be given a precisely defined goal (or rather goals: it is more usual that there are multiple choices to be made). A common, but not universal, attribute of these technologies is their use of data[7]. The advances and benefits created by electricity and railways did not arise by virtue of

[2] See the analysis of "collective technologies" in Dana Sanchez, *Collective technologies: autonomous vehicles*, ACOLA November 2014. https://acola.org.au/wp/PDF/SAF05/2Collective%20technologies.pdf

[3] There is a huge literature on GPTs. Out of the many choices I pick just a few:
Paul A. David and Gavin Wright, General Purpose technologies and Surges in Productivity: Historical Reflection on the Future of the ICT revolution, presented to the International Symposium on *Economic Challenges of the 21st Century in Historical Perspective*, Oxford, 2-4th July 1999;
Timothy Bresnahan, *General Purpose* technologies Chapter 18 of Handbooks in Economics Volume 2, Elsevier 2010.
Richard G. Lipsey, Kenneth I. Carlaw and Clifford T. Bekar, *Economic Transformations: General Purpose Technologies and Long Term Economic Growth*, Oxford University press, 2005.
I made considerable use of the GPT concept in the *Technology and Australia's Future Report* (op. cit.)

[4] Paul A. David, Digital Technology and the Productivity Paradox: After Ten Years, What has been Learned? Presented at *Conference on Understanding the Digital Economy: Data, Tools and Research*, U.S. Department of Commerce, May 1999
The issue of economic indicators being poorly designed to capture the productivity growth of new GPTs is covered on pages 72ff of Nathan Rosenberg, *Schumpeter and the Endogeneity of Technology*, Routledge 2000.
On the long reach of old technologies: David Edgerton, *The Shock of the Old: Technology and Global History since 1900,* Oxford 2007.
On the general dynamics of technology and society: Joel Mokyr, *The Gifts of Athena: Historical Origins of the Knowledge Economy,* Princeton University Press, (2002)
Dominque Foray and Christopher Freeman, *Technology and the Wealth of Nations: The Dynamics of Constructed Advantage*, OECD 1993.
Chris Freeman and Francisco Louca, *As Time Goes by: From the Industrial Revolutions to the Information Revolution*, Oxford University press 2001.
David S. Landes, *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*: Cambridge University Press, (2003)

[5] The phrase "technology as useful knowledge" is relied upon by Mokyr, who claims he did not invent it (Joel Mokyr, *The Lever of Riches: Technological Creativity and Economic Progress*, Oxford University Press, (1990)). It was explicitly analyzed by Edwin T. Layton Jr., Technology as Knowledge, *Technology and Culture* 15(1), 31-41 (1974). See also Marc J de Vries, The Nature of Technological Knowledge: Extending Empirically Informed Studies into What Engineers Know, *Techne* 6(3), 1-21, 2003.

[6] National Resources Committee (Subcommittee on Technology), *Technological trends and national policy including the social implications of new inventions*. Washington: National Resources Committee, (1937).
See also section 3.6.2 of *Technology and Australia's Future* (op. cit.).

[7] See Kirsty Douglas, *Technologies for Data*, ACOLA 2015,
https://acola.org.au/wp/PDF/SAF05/11Technologies%20for%20data.pdf

people sitting around discussing better and better definitions of what really is electricity, or what is the essence of a railroad!

It is also helpful to consider the social problems that the technology seeks to solve; it is never a single problem, and there is also never agreement on whether it has been solved. For railways this seemed clear at first, but it gets murky. Likewise, with electricity. Paul David observed that when electrical motors were first introduced into factories, they were used to replace the large steam engines. The labyrinth of belts and pulleys was kept. The next generation had small individual motors per machine. Then it was realised that the entire layout of factories could be changed (and was). This was not foreseen, nor foreseeable!

Regarding the problems that AI is used to solve, many of them (not all) are problems of control, and it is useful to conceive of the future in the context of the control technologies used in the past. Paramount of these, and the one crying out for regulational intervention, is advertising[8].

But, it will be argued, this misses the point: surely AI is different? Surely, you cannot understand it using old-fashioned views of technology?! Yes and no. It unquestionably is a technology, and shares many features of other technologies. The stand-out distinction, which is hardly new, was clearly identified over half a century ago:

> We are now coming to realize that man and the machines he creates are continuous and that the same conceptual schemes, for example, that help explain the workings of his brain also explain the workings of a "thinking machine. and his refusal to acknowledge this continuity, is the subs which the distrust of technology and an industrialized society has been reared. Ultimately, I believe, this last rests on man's refusal to understand and accept his own nature-as a being continuous with the tools and machines he constructs[9].

These anxieties were prevalent in the development of automation in the second half of the twentieth century. The concerns, especially with regards to employment, were the major driver for the Myer's report on technological change[10]. I recommend that with all of these concerns, that there is much we can learn from the past!

Technological Prediction

> *Over the past year, with the turn of the new century – and Millennium – the media have been filled with speculations. You might call it the 'Where is technology taking us?' syndrome. I want to assert in the strongest possible terms what I regard as the only possible serious answer to that momentous question: We don't know. In fact, I believe that we can't know*[11]*.*
>
> *Viewed in retrospect, the evolutionary path of a fully developed general-purpose technology has the appearance of inevitability. When the technology is in its infancy, however, an observer looking into the future cannot conceivably know if it will turn out to be a modest advance operating over a limited range, a general purpose technologies, or anything in between*[12]*.*

[8] James R. Beniger, The Control Revolution: Technological and Economic Origins of the Information Society, Harvard University press 1986.

[9] Bruce Mazlish, The Fourth Discontinuity, *Technology and Culture*, 8(1), 1-15 (1967).

[10] Committee of Inquiry into Technological Change in Australia, *Technological change in Australia, Vols 1-4*. Canberra, (1980). For criticism of this report see footnote 5, page 7 of *Technology and Australia's Future* (op. cit.)

[11] Nathan Rosenberg, *Challenges to the social sciences in the new millennium*. Paris: OECD. 7-24 p, (2001)

[12] Richard G Lipsey, Cliff Bekar, Kenneth Carlaw, What requires explanation? In: E. Helpman, editor. *General purpose technologies and economic growth*. Cambridge: MA: MIT Press. pp. 14-54 (1998); page 48.

> *One "lesson", however, to be drawn here is that any social invention is bound to be surrounded by prophecies, pro and con. Their value is usually very little, expect as statements of mythical aspiration…*[13]

The first two questions are how will the technology advance, and what will its impacts be? Much prognostication with regard to AI seems to forget that AI is still a technology, and as wonderous and new as it seems (and indeed is) *it is no more radical, disruptive or transformational than previous GPTs.* Such a comparison of the new with the old is humbling and provides much-need context. MIT historian Bruce Mazlish admirably made such a comparison (between railroads and the space-program) in the 1960s which was as breathless about space-technologies as we seem to be about AI (ironically, some 50 years later, Australia has now jumped on the space bandwagon!)[14]. Study of the degree to which impacts could be foreseen, or even understood as they happened[15], with regard to past technologies will provide useful counterweight to the notion that everything about AI is different. It simply is not.

The temptation to ask the questions of prediction regarding technology is indeed great. But it is prudent to ask whether such questions are in fact answerable with any useful accuracy. Fortunately, *that* meta-question is answerable, by the simple device of looking at such predictions made long ago, and comparing their prediction with what has actually come to pass. The answer to this meta-question is unequivocal: *one can neither accurately predict particular technological advances, nor the successful applications of the technology*. Neither asking the inventors[16], polling "experts"[17], appeal to the purported superior prognostic ability of science fiction authors[18], the structured construction of future "scenarios"[19], mechanical "horizon scanning" processes[20], or citation or patent analysis[21] empirically succeed; indeed their track-record is appallingly bad – a monkey with a coin will typically do better. And schemes such as Delphi which purport to have a magical method of aggregating the views of many improve the situation; whilst they generate consensus, they do not lead to truth[22]. The same holds for fancy mathematical modelling[23]. The answer is surprising, and (personal experience shows) is often aggressively rejected. The surprise comes from the fact that certain "hypotheses of technological progress"[24] (the most famous being Moore's law) appear to have an incredibly good predictive validity. How can one reconcile these two points, and what does it mean in respect of the AI prediction questions?

When one looks at the evidence, the only really surprising thing is that the hypotheses of technological progress work at all, given the failure of other predictions. There are several possible explanations, but the honest truth is nobody knows – we await the development of a rigorous statistical mechanics for evolving technological systems! In one case, an answer is readily available:

---

[13] Bruce Mazlish, Historical Analogy: The Railroad and the Space Program and Their Impact on Society, pages 1-53 in Bruce Mazlish (Ed), *The Railroad and the Space Program: An Exploration in Historical Analogy*, MIT Press 1965, Page 37.
[14] Ibid.
[15] Astonishingly, it is even hard to ascertain the impacts *after the fact* – witness Robert Fogel's masterly historical analysis (which won him his Nobel prize in economics) on the societal impact of the railroad in the American West (executive summary: if the railroad had not come along, in aggregate things would be little different, with the change in GDP in the late 19th century down by only 1%; the system of canals would likely have been extended and served much of the purpose of the railroad (at least for freight movements, which is what drove the economic exploitation of the American west). (Robert Fogel, *Railroads and American Growth: Essays in Econometric History*, The John Hopkins University Press, 1964).
[16] See section 3.3.1 of *Technology and Australia's Future* (op. cit.)
[17] See section 3.3.2 of *Technology and Australia's Future* (op. cit.)
[18] See section 3.3.3 of *Technology and Australia's Future* (op. cit.)
[19] See section 3.3.4 of *Technology and Australia's Future* (op. cit.)
[20] See section 3.3.5 of *Technology and Australia's Future* (op. cit.)
[21] See section 3.3.6 of *Technology and Australia's Future* (op. cit.)
[22] See section 3.4 of *Technology and Australia's Future* (op. cit.)
[23] Ibid.
[24] See section 3.3.7 of *Technology and Australia's Future* (op. cit.)

self-fulfilling prophecy – Moore's law has become and industry roadmap that the semiconductor industry is committed to see fulfilled. But even without such effects, trends such as unit costs of Photovoltaic power generation technologies tend to track a trend line. The thing to realise is what is being predicted here is not the specific technologies that bring the price down – almost always there is a large complicated mess of technologies. Rather all one is seeing is an aggregation – the law of large numbers in action as it were. And one should be very wary of presuming the ongoing continuation of current trend lines.

Regarding prediction of social impact (rather than technological advances per se), the track record is even worse, especially when technologists make the predictions. New technologies, like other innovations *diffuse* through society in hard to fathom manner[25]. The best bet (following Gilfillan and Ogburn) is to focus on the solution to societal problems. I predict that the class of technologies know as AI will contribute to the partial solution of all the problems society will face in the future. For almost none of them will it be the sole solution. The solutions will be imperfect. There will be technological failures. When enough people are harmed, the uses of the new technology will be more stringently regulated. And the new technology will create entirely new problems (or more likely, amplify existing ones) to a degree that is not foreseen in advance.

The one positive use of technological prediction in the narrow sense that I embraced in the 2015 ACOLA report[26] was the sense in which envisaging the technological future helped one invent it ("the best way to predict the future is to invent it."). To that end I summarise the essence of what I think are the big opportunities in terms of improving the technology of machine learning in particular:

What needs inventing

In the spirit of Alan Kay's "the best way to predict the future is to invent it" I list below four grand problems in Machine Learning (ML) that I think are important in terms of research (in the precise sense that these are the problems I am working on or plan to work on).

Problem-oriented foundations for ML

The majority of machine learning research is "technique oriented". There is compelling evidence that a problem-oriented approach is more powerful and would resolve many issues[27]. The root cause of the problem is that whilst AI is based on computer-science, and computer science is based on algorithms, there is no formal notion of what an algorithm actually is[28]. The challenge then is to formally describe the relationship between machine learning *problems*, in the same manner that mature engineering disciplines can do. Techniques come and go, but problems tend to be more long-lasting, and, more to the point, it is the problems that people care about.

[25] Everett M. Rogers, *Diffusion of Innovations*, (Fifth edition), Free press 2003.
[26] See section 3.3.1 of *Technology and Australia's Future*, op. cit.
[27] John R. Platt, Strong Inference: Certain systematic methods of scientific thinking may produce more rapid progress than others, *Science* 146(3642), 347-353, (16 October 1964).
[28] This fact is undeniable (you will search in vain through the 1292 pages of [Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, *Introduction to Algorithms* (3rd Edition), MIT press, 2009] for a formal definition);. For a formal attack on the problem, see Andreas Blass and Yuri Gurevich, Algorithms: A quest for absolute definitions, *Bulletin of the EATCS* 81 195-225, (2003); Andreas Blass, Nachum Dershowitz, and Yuri Gurevich. When are two algorithms the same? *Bulletin of Symbolic Logic* 15, no. 2 145-168, (2009). Many computer scientists just go into denial over the matter, and machine learning researchers, on the whole, seem to love nothing more than a new algorithm. There is some progress, but my point still holds, especially with regard to machine learning problems, taking algorithms as the fundamental atoms of inquiry is not the best bet.

Fundamental Limits of ML

Once one has a clear story about the range of machine learning problems, and their interrelationship, one can ask the subsequent question "how well can they be solved". This includes specific questions such as what characterises the difficulty of a machine learning task? What are the relationships between the difficulty of a statistical decision problem and the limits of physics (via Landauer's principle for example). What are the fundamental information-theoretic limits of privacy and fairness in ML when applied to data about people? What are the fundamental trade-offs between (for example) fairness and performance?

Ethical foundations for ML

There is now considerable interest, and rapidly growing work on the ethical issues of machine learning. Like the rest of the field (*vide supra*) this tends to be focussed greatly on algorithms. There is a much broader set of questions concerning suitable ethical frameworks to use to even state the problem of ethical machine learning. There will not be a single answer, so the immediate question then becomes what are the relationships between, and the implications arising from different choices of framework. I believe this line of work will not only illuminate and progress the science and technology of machine learning, but offers a significant opportunity to advance the state-of-the-art in ethics and moral philosophy. The question is necessarily highly interdisciplinary and will be pursued via interdisciplinary collaboration.

End-to-end trust for ML

Trust is central to the social acceptability of a technology, and AI is no different. I believe that focussing on all aspects of trust in AI is perhaps the most important thing future research should concentrate on. The *Mnemosyne* proposal, which I have already shared with this ACOLA group, amplifies this argument, and spells out a detailed list of specific technical problems to be considered.

Managing Technological Change

The third question could be profitably reframed as: *How can Australia facilitate and derive maximal benefit from the technological revolution triggered by the latest GPT, namely AI?* I submit that the best frame within to consider this question is that of evolution[29]. The evolutionary perspective is rich, robust, and well corroborated, embraces uncertainty[30], and it helps pivot away from unanswerable questions (such as those of prediction) to those of facilitation – how can we speed up evolution (modularity and interoperability, and avoidance of categorical lock-in)? What qualitative patterns should we look out for? (punctuated equilibria and other non-ergodic phenomena).

Perhaps the most important distinction to make is that between a technology and its use. The broad consensus, reflected in practice, is that regulation of technology is best done by its use. This makes

[29] John Ziman, *Technological Innovation as an Evolutionary Process*, Cambridge University Press, (2000)
Patrick Kelly, Melvin Kranzberg, Frederick A. Rossini, Norman R. Baker, Fred A. Tarpley, Jr and Morris Mitzner, *Technological Innovation: A Critical Review of Current Knowledge* (Volume 1, The Ecology of Innovation and Volume 2, Aspects of Technological Innovation), Advanced Technology and Science Studies Group, Georgia Tech, Atlanta, Georgia, (February 1975)
George Basalla, *The Evolution of Technology,* Cambridge, Cambridge University Press, (1988)
Giovanni Dosi, Technological Paradigms and Technological Trajectories. *Research Policy,* 11, 147-162, (1982)
Vernon W. Ruttan, *Technology, Growth and Development: An Induced Innovation Perspective,* Oxford University Press, (2001)
Paul A. David, Clio and the economics of QWERTY, *The American Economic Review*, 75(2), 332-337, (May 1985)
Frank W. Geels, *Technological Transitions and System Innovations: A co-evolutionary and Socio-Technical Analysis*, Edward Elgar, Cheltenham, (2005).

[30] Nathan Rosenberg, "Uncertainty and Technological Change," pages 17-24 in Dale Neef, Anthony Seisfeld and Jacquelyn Cefola (Eds), *The Economic Impact of Knowledge*, Butterworth Heinemann 1998.

sense since it is rare that technologies themselves are the cause of harm; it is how they are used: there are no technologies that are wholly good, or wholly bad[31].

The general question of what one might do in face of unpredictable evolutionary technological change is complex. Perhaps the categorization of problems in chapters 6 and 7 of *Technology and Australia's Future* will be of help here. First one should understand how to *evaluate* the technology. A common error (with all new technologies) is to accept (as just the way the world is) the downsides of existing technologies, but to overreact to perceived downsides of new technologies. This pattern is so pervasive, it may well be called a law. As a concrete example, concerning machine learning is the use of ML decision making regarding people – what is often called 'algorithmic decision making.' There is no doubt that there are many ways this can cause harm, or provide poor outcomes. A reasonable comparison however is to compare it to human decision making. There are many studies showing that mathematical formulae (which is what algorithms actually compute) perform better than intuitive decision making by people[32].

Second one needs to categorise the types of interventions possible. In Chapter 7 of *Technology and Australia's Future*, we singled out personal behaviour (tinkering), education for adaptability, attitudes to failure (individual and institutional), experimentation versus traditional approaches to policy setting, interoperability, modularity and standards (which I note you already have in your table of contents), regulation (ditto), and government investment in technological research and development. I suggest you consider each of these categories with respect to AI.

The evolutionary model of technological change posits that technological changes are the results of lots of little changes. These are typically mediated by employee mobility[33]. The economists refer to this as "learning by doing." One implication is the particular importance of a highly skilled and mobile workforce. Indeed, one of the key motivations for the creation of NICTA[34] was the production of such a skilled workforce, and at its peak it was graduating around 100 PhDs per year, mostly in Artificial Intelligence.

It is relevant to focus upon institutions like NICTA (whereas Australia closed its AI powerhouse down, most other Western countries are busy building them up). The reason is that the nature of AI *is* different to the traditional science / engineering split (although I think that that split is likely a poor model for the reality elsewhere too). The one thing I draw attention to is the platform nature of the

---

[31] Kranzberg's law adds a nuance: technology is neither good nor bad; neither is it neutral. See Melvyn Kranzberg, Technology and History: "Kranzberg's Laws", *Technology and Culture* 27(3), 544-560, 1986. His second law is equally apposite for the present purpose, signalling the rich pathways in which new technologies arise: "invention is the mother of necessity." We met his third law already indirectly in the earlier mention of collective technologies: "Technology comes in packages, big and small."

[32] An early example is Robyn M. Dawes, The Robust beauty of Improper Linear Models in Decision Making, *American Psychologist* 34(7), 571-582, July 1979 (which showed that even outrageously simple mathematical models can outperform people). A slightly more recent popular article is John A. Swets, Robyn M. Dawes and John Monahan, Better decisions through science, *Scientific American* October 2000.

[33] Bruce Fallick, Charles A. Fleischmann and James B. Rebitzer, *Job hopping in silicon valley: some evidence concerning the micro-foundations of a high technology cluster*, NBER working paper 11710, October 2005.

[34] NICTA was funded by the federal government until 2015. There was never any reason given for the cessation of funding. Its performance was outstanding (20 times the startups produced per federal research dollar; twice the number of papers per scientist per year than CSIRO, and 18 times the number of PhD students supervised per scientist); see the NICTA submission to the Boosting the Commercial Returns of Research enquiry for evidence and references: https://submissions.education.gov.au/Forms/higher-education-research/_layouts/SP.Submissions/ViewDoc.ashx?id=%7Bfa031744-1c0f-4a39-a45a-e42b6fd085a1%7D . In its final year, NICTA received $42m from the federal government. At least two thirds of its research was in the core areas of artificial intelligence. The significant global position Australia had in 2014 is now largely lost. (The machine learning group, which I ran, was rated by independent international review as amongst the top five in the world; my estimate now is that no group in Australia is in the top 50 in AI).

technology. "Platform" can refer to platform businesses (which indeed are enabled by AI) or the general purpose and generative nature of the technology itself. With regard to the latter, I think it is important to distinguish between the wealth that new technologies generate, and how it is appropriated[35]. I believe the NICTA experiment showed the value, with regard to federally funded support for AI, of *not* attempting to appropriate the returns, but rather to focus on generating wealth for the country. This is particularly pertinent for general purpose technologies, such as AI. This is tied intimately to notions of open innovation, and I can really do nothing more here than mention it.

Finally, and related to the previous point, the lessons drawn in *How the West Grew Rich*[36] are very apposite with regard to AI – big innovations come from small groups with minimal oversight. If you try to plan and predict and control things too much, you will kill the goose that lays the golden egg.

Comments on Selected parts of the Draft Table of Contents

I was provided with the draft table of contents for the report ACOLA is preparing. Noting the immense difficulty of inferring what might be envisaged for the subheadings in the draft table of contents, I nevertheless offer a few points that may be of interest. I have focussed on points that might be differentiated from the consensus view. In the interests of brevity, I have been pretty telegraphic here; I am happy to expand on any of these points if anybody reads them and is actually interested…

2.0 Technology and Applications

The very heading here implies a old-fashioned linear model[37] of technological development and change, that is widely refuted by the facts[38]. I think it would be worthwhile understanding, and documenting, the complex mechanisms at play in the development of new technologies, especially GPTs. The GPT viewpoint has been widely used by economists. It has, unfathomably, not been widely used in all the other dimensions of technology[39].

Related to the above is the boundary between technology and humans. Problematic with many technologies, with AI, this notion of a simple dichotomy would appear to do more harm than good; I think you need to grapple with the central issue of boundary between human and machine.[40]

---

[35] The issue of appropriability was raised an astonishingly long time ago, but is equally astonishingly absent from the vast majority of government policy interventions in the half a century since. See Richard R. Nelson, The Simple Economics of Basic Scientific Research*, The Journal of Political Economy* 67(3), 297-306, 1959. See recent work: Richard C. Levin, Alvin K. Klevorick, Richard R. Nelson and Sidney G. Winter, "Appropriating the returns from industrial research and development," *Brookings papers on economic activity* 3, 783-831, 1987; Emmanuel Dechanaux, Brent Goldfarb, Scott Shane, Marie Thursby, "Appropriability and Commercialization: Evidence from MIT Inventions", *Management Science* 54(5), 893-906 (2008); Katherine A. Hoye, *University Intellectual Property Policies and University-Industry Technology Transfer in Canada*, PhD thesis, University of Waterloo, 2006.
See especially Giovanni Dosi, Patrick Llerena, and Maoro Sylos Labini, "The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called 'European Paradox'," *Research Policy* 35, 1450-1464, (2006).

[36] Rosenberg, Nathan, and Birdzell L.E. Jr. *How the West grew rich: The economic transformation of the industrial world*. Basic books, 2008. See also the submission referenced in footnote 34.

[37] Whilst widely discussed, and ironically still acted upon, it seems the model *never* existed: David Edgerton, 'The linear model' did not exist: reflections on the history and historiography of science and research in industry in the twentieth century, in Karl Grandin and Nina Wormbs (eds), *The Science-Industry Nexus: History, Policy, Implications*, Watson, 2004.

[38] See the discussion, with extensive references in chapter 2 (the shaping of technology) in *Technology and Australia's Future* op. cit.

[39] The one exception, and well worth looking at is Terry Shinn, New sources of radical innovation: research-technologies, transversality and distributed learning in a post-industrial order, *Social Science Information* 44, 731-764, 2005. See especially pages 751ff.

[40] Bruce Mazlish, *The Fourth Discontinuity,* op. cit.

2.1 Next decade of technology and infrastructure requirements to support it

I suggest you think hard about what is meant by infrastructure. In particular, do not be seduced by big machines, and think about the entire software stack, and how it can be developed and maintained in an open manner (so it is not appropriated by individual corporates)[41].

3.1 Employment and the workforce

Much has been written on AI and employment. I summarised a considerable literature a few years ago[42]. I note that the anxieties we have now, are a replay of those in the 50s and 70s. It is not clear there is a fundamental difference. Which is not to say there is not a problem. Much is made of replacement of "routine" jobs, but routine is a matter of degree. Interestingly the distinction is tied to that of the distinction between a tool and a machine[43]. Lewis Mumford's famous distinction between them is as follows:

> The essential difference between a machine and a tool lies in the degree of independence in the operation from the skill and motive power of the operator: the tool lends itself to manipulation, the machine to automatic action. …. The difference between tools and machines lies primarily in the degree of automation they have reached…[44]

Rather than autonomy per se, perhaps it is more helpful to view the question from the perspective of "conviviality"[45]. The fact that Illich's *Tools for Conviviality* played such a significant role in the development of the personal computer is surely apposite.

I trust you folks are already aware of it, but especially in the context of jobs, I strongly encourage you to look at Rod Brook's admirable blog post on the seven deadly sins of predicting the future of AI[46].

3.2 Education, skills and training

Everything we said in sections 7.3-7.5 of *technology and Australia's Future* is very relevant here[47]. Viewing technology as knowledge, and accepting the complex dynamic of technological change, I

---

[41] See Ed Lazowska, "A Plea for Greater Attention to Data-Intensive Discovery, Greater Investment in Intellectual and Software Infrastructure, and Greater Use of the Commercial Cloud" (Presentation to the NRC Computer Science & Telecommunications Board Colloquium on Future Cyberinfrastructure for Scientific Discovery, September 2016), http://lazowska.cs.washington.edu/Cyberinfrastructure.pdf who stresses the importance of software infrastructure. See also Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor: Deep Blue. http://hdl.handle.net/2027.42/97552 and Nadia Eghbal, *Roads and bridges: The Unseen Labor Behind our Digital Infrastructure*, Ford Foundation, July 2016, https://www.fordfoundation.org/library/reports-and-studies/roads-and-bridges-the-unseen-labor-behind-our-digital-infrastructure/ . For those who have not seen it before, it is also worth reading Susan Leigh Star's "the Ethnography of Infrastructure", *American Behavioural Scientist* 43(3), 377-391, 1999, for a compelling argument as to why infrastructure needs to be thought of much more broadly than is usually done, especially in respect of its social dynamics.

[42] Michelle Nic Raghnaill and Robert C. Williamson, *Technology and Work*, ACOLA 2015, https://acola.org.au/wp/PDF/SAF05/7Technology%20and%20work.pdf

[43] Miguel Barroso Morin, *General Purpose Technologies: Engines of Change?* PhD Thesis, Columbia University 2014.

[44] Lewis Mumford, *Technics and Civilisation*, George Routledge and Sons, 1946. (Page 10). Re tools versus machines, see also pages 41ff of Branden Hookway, *Interface*, MIT press, 2014.

[45] Confer Ivan Illich, *Tools for Conviviality*, Fontana 1973, which was a singnificant inspiration behind the invention of the personal computer (see *Lee Felsenstein and the Convivial Computer*, (July 2007), http://conviviality.ouvaton.org/spip.php?article39). The notion of conviviality has been revived (Dan McQuillan, Algorithmic paranoia and the convivial alternative. *Big Data & Society.* 2016 Nov;3(2): 2053951716671340) and would serve as a very useful frame for the current (legitimate!) anxieties regarding the over-reaching control of commercial platforms that exploit AI technologies.

[46] https://rodneybrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/

[47] See also the paper on tinkering by Kat Jungnickel, *Tinkering with technology: Examining Past practices and imagined futures*, ACOLA 2015, https://acola.org.au/wp/PDF/SAF05/9Tinkering%20with%20technology.pdf

also think questions of user-driven innovation are relevant here[48]. Finally, I note that the skills needed to make use of sophisticated new technologies are essentially the same as those needed to invent it in the first place. So, a policy of letting folks overseas invent it, and merely applying it here simply will not work.

3.3 Society and the Individual

See chapter 5 of *Technology and Australia's Future* on the meaning of technology[49], which is complex, contradictory, and seems to have a much greater effect on adoption than many technologists would suspect. It is also something amenable to manipulation.

4 Considerations for the development and implementation of AI

How technologies are evaluated makes a big difference to their use. I think the point is especially pertinent with something like AI (given its somewhat unusual characteristics). I would look carefully at all the problems that arise in the evaluation of technology generally (Chapter 6 of *Technology and Australia's Future)* and consider especially their significance for AI.

4.1 Trust, transparency and fairness

Trust is indeed central to the social acceptability of all technologies. You might find it helpful to look at some of what I wrote regarding trust in the DATA61 science vision[50].

4.1.1 Trading off accuracy and fairness

Indeed it is true that if you impose an additional constraint, accuracy can only get worse[51]. Much of the literature on the topic has a tendency to come across as social justice warriors in action. For example, there was a lot of press about COMPAS algorithm for recidivism prediction. This is a good case to consider. One can concern oneself with the fairness to those incarcerated, regarding whether they receive parole. But such considerations ignore the fairness concern regarding the potential future victims who they may harm. If one distinguished on the basis of the race of the potential future victims, then given the structure of the situation, it could very easily be the case that increased "fairness" to the internees results in decreased fairness to their spouses (many people incarcerated are in jail because of domestic violence). My point is simply this: there is no universal and "correct" notion of what is fair. One needs to always keep this in mind when considering trade-offs.

Related to fairness, and something you should address is *transparency*. There is a nice recent paper arguing why transparency is *not* really the solution many might thing it is[52].

4.1.2 Autopilot phenomenon.

I am guessing what is mean by the autopilot problem, but I suspect it is what has gone under the

[48] See the works of Eric von Hippel.

[49] And Kat Jungnickel, From Frankenstein to the Roomba: The changing nature and socio-cultural meanings of robots and automation, ACOLA 2015, https://acola.org.au/wp/PDF/SAF05/3From%20Frankenstein%20to%20the%20Roomba.pdf

[50] Robert C. Williamson, *Data you can trust: technology that works for you: DATA61's Science Vision*, 2016, https://www.data61.csiro.au/en/who-we-are/our-science-vision

[51] One can mathematically formulate this problem. Interestingly, the trade-off is not dependent so much on any algorithm, or any technology at all – it is an intrinsic function of the underlying data. I think this has significant implications for how one thinks about this tension. See Aditya K. Menon and Robert C. Williamson, The cost of fairness in binary classification. In Conference on Fairness, Accountability and Transparency 2018 Jan 21 (pp. 107-118) http://proceedings.mlr.press/v81/menon18a.html See especially the non-mathematical discussion in Appendix H of the supplementary materials.

[52] Kroll, Joshua A. and Huey, Joanna and Barocas, Solon and Felten, Edward W. and Reidenberg, Joel R. and Robinson, David G. and Yu, Harlan, Accountable Algorithms, *University of Pennsylvania Law Review* 165, 633-705, 2017.

name of the ironies of automation for 35 years[53], and connection to that earlier literature would be helpful (to signal to your readers that this is hardly a new phenomenon, and that there are well developed ways of managing it).

4.1.3 Social licence

Regarding the GM backlash, see the working paper by Dana Sanchez[54]. Regarding their use in monopolistic platform-based businesses, I commend Edward Tenner's latest book[55].

4.1.4 Public communication

I think advertising fits under this heading, and it is great thing to focus upon. I am astonished at the equanimity with which most people accept the premise of the advertising industry. Robert McChesney calls it "the greatest concerted attempt at psychological manipulation in all of human history."[56] [57].

Advertising is "a form of communication intended to persuade an audience (viewers, readers or listeners) to purchase or take some action upon products, ideals, or services"[58]. The key point here is that the intent in advertising is to *persuade*, not to merely *provide information*. Advertising does not just give people what they want, even if it is hyper-personalized. The purpose of advertising is to give the *advertiser* what they want. Its value to businesses that do advertise is that it persuades people to buy their products and services.

> As database companies and technologies get better at targeting, data crunching, and understanding behaviors—and addressable advertising can deliver on its promises—inevitably the consumer will find it creepy and we will experience a backlash, Jacobs believes. "We've already seen this with social networking privacy policies," he says. "That will be the death of a lot of companies that offer behavioral targeting and sit somewhere in the middle of the value chain between publishers and advertisers."[59]

A careful study[60] carried out by the University of Pennsylvania and the University of California at Berkeley found that "Contrary to what many marketers claim, most adult Americans (66%) do not

---

[53] Lisanne Bainbridge, Ironies of Automation, *Automatica* 19(6), 775-779 (1983).

[54] Dana Sanchez, *Genetically modified crops: how attitudes to new technology influence adoption*, ACOLA, March 2015, https://acola.org.au/wp/PDF/SAF05/4Genetically%20modified%20crops.pdf

[55] Edward Tenner, *The Efficiency Paradox: What Big Data Can't Do*, Alfred A. Knopf, 2018.

[56] *The Political Economy of Media: Enduring Issues, Emerging Dilemmas*, Monthly Review press, May 1, 2008), p. 277.

[57] Kalle Lasn describes it as "the most prevalent and toxic of the mental pollutants. From the moment your radio alarm sounds in the morning to the wee hours of late-night TV microjolts of commercial pollution flood into your brain at the rate of around 3,000 marketing messages per day. Every day an estimated twelve billion display ads, 3 million radio commercials and more than 200,000 television commercials are dumped into North America's collective unconscious". (*Culture Jam: The Uncooling of America*, William Morrow & Company, (November 1999))

[58] Advertising (Wikipedia)

[59] KPMG, quoting Walker Jacobs, the Senior Vice president in charge of online advertising sales at Turner Sports and Entertainment. From pages 26-27 of Networked Advertising: Growing Revenue in a Highly Fragmented business , 20 April, 2010. The report concludes: "The debate about how best to monetize online and mobile content is reaching a fever pitch. Some media moguls such as Rupert Murdoch have seemingly decided that advertising revenue is simply not enough and are now instituting "paywalls" on their Web sites to extract payment for newspaper content." Interestingly this change provides some data regarding how many people are willing to pay: "There has been a 60 per cent drop in traffic to The Times website over the past couple of weeks since the introduction of its registration wall." See The Times begins charging for online access, Internet Advertising Bureau, 5 July 2010. It would be useful to get statistics of how many people are willing to pay the £2/week that is being charged. See the Times' own Why do you believe that customers will pay when they can get news elsewhere on the web for free? (These references are admittedly dated; they are drawn from a (confidential) report I wrote for Microsoft in 2010; but I think they are still pertinent)

[60] Joseph Turow, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley and Michael Hennessy, *Contrary to what marketers say, Americans Reject Tailored Advertising and three activities that enable it*, University of Pennsylvania and University of California, Berkeley, September 2009.The detailed analysis in this 27 page report (based on a survey of 1000 people) cannot be done justice here. Their conclusion was prescient: "Companies need to respect their publics rather than to treat

want marketers to tailor advertisements to their interests." The authors said, "It is hard to escape the conclusion that our survey is tapping into a deep concern by Americans that marketers' tailoring of ads for them and various forms of tracking that informs those personalizations are wrong." Interestingly the authors also concluded that

> [C]ontrary to consistent assertions of marketers, young adults have as strong an aversion to being followed across websites and offline (for example, in stores) as do older adults. 86% of young adults say they don't want tailored advertising if it is the result of following their behaviour on websites other than one they are visiting, and 90% of them reject it if it is the result of following what they do offline.

Recently, in the context of pondering the future of the company, Google's CEO has recently made the astonishing claim that

> I actually think most people don't want Google to answer their questions. They want Google to tell them what they should be doing next.[61]

The issue of AI and advertising has come to the attention of some recently with the Cambridge Analytica (CA) scandal. The story of how data was exfiltrated from Facebook is not the real issue though. The problem is that even when Facebook keeps all personal data secure, it can still be used to manipulate people. Until the CA scandal, Facebook actually bragged of its ability to do so![62] I would love to see some material in your report about this weaponization of friendship. It is readily regulatable. It is not a flaw in the technology – merely on how it is used (and abused).

4.2 Regulation and Regulatory Frameworks

I think this is particularly important for AI. I suspect that the general rule, that applies close to universally, that technologies are best regulated by their use, will apply also to AI. See section 7.8 of *Technology and Australia's Future* for detailed citations to the significant literature on this topic.

4.2.4 Appealing algorithmic decisions

I am guessing that the phrase 'algorithmic decisions' is used, as is common, to contrast decisions made by an AI-based computer program, versus those made my people. Ironically, many (most?) decisions made about individuals prior to the use of AI were entirely algorithmic: the very essence of bureaucracy is "calculable rules"[63]. As in the past, the way to handle the possibility of things going awry, is layers of due process.[64] The concerns that many (legitimately) have regarding such decisions based on data and codified rules are not new, and there is value in understanding how these concerns were dealt with in the past[65]. The use of computerised AI technologies actually offers the opportunity to implement much better and more rigorous due process, because a computer can be

---

them as objects from which they can take information in order to optimally persuade them with no clear option not to participate." (page 26).

[61] See Holman W. Jenkins Jr, Google and the Search for the Future, The Wall Street Journal, 14 August 2010
Schmidt went on to say: "The power of individual targeting—the technology will be so good it will be very hard for people to watch or consume something that has not in some sense been tailored for them." A reader's (Stephen Rollins) comment makes the point succinctly: "if any SALESPLATFORM tells me I need something, I mostly know I don't."

[62] Facebook hid previously available govt and politics ad section https://theintercept.com/2018/03/14/facebook-election-meddling/ ; see also https://www.theguardian.com/uk-news/2018/mar/21/why-have-we-given-up-our-privacy-to-facebook-and-other-sites-so-willingly "Facebook also boasts to advertisers about how much it knows about its users – and how effective it can be at influencing their minds"

[63] Bureaucratization offers above all the optimum possibility for carrying through the principle of specializing administrative functions according to purely objective considerations. Individual performance are allocated to functionaries who have specialized training and who by constant practice increase their expertise. "Objective" discharge of business primarily means a discharge of business according to **calculable rules** and "without regard for persons." Max Weber, *Economy and Society: An Outline of Interpretative Sociology*, University of California Press, 1978, page 975. There is a vast literature following on from Weber. See for example Frederick Schauer, *Playing by the Rules: A Philosophical Examination of Rule-based Decision-making in Law and Life*, Clarendon Press, 1991;

[64] Daniell Keats Citron, Technological Due Process, *Washington University Law Review*, 85(6), 1249-1313 (2008).

[65] Alain Desrosieres, *The Politics of Large Numbers: A History of Statistical Reasoning*, Harvard University Press, 1998.

programming to *always* record the precise reasons for any decision, and such records are thus available to audit[66].

4.4 Ethical frameworks

The general question of what are suitable ethical frameworks is a fine research question – precisely because at present I do not think there is an answer. And there will not be a single answer, but many.

4.4.2 Privacy and surveillance

This is a significant problem area. The difficulties primarily arise (as usual) with the use to which technologies are put. I trust you will be looking broadly at notions of privacy, including for example Helen Nissenbaum's contextual integrity[67], which also focussed on the use to which data is put. The interesting thing is that business practices (that are effectively technologies of control such as advertising, have had the roots of the moral problem baked in for years[68]. It is only now that advertising can be done in such a fine-grained manner that its harm is manifest. But to be clear, the problem is with the toxic business of advertising, not the AI that allows corporations like Facebook and Google to serve up the sludge[69]. Perhaps a better word to use is "trumpery"[70]. Richer views of privacy, especially privacy as *autonomy* are not new[71], but they are (I think) exactly what is needed more than ever, and I encourage you to weave this into your report.

4.4.3 Ethical frameworks

The heading could mean anything. I hope you point out that *any* ethical assessment of technology presumes an ethical framework, and that there are many to choose from, and generally people disagree. This sounds like a triviality, but when you see people talking about "bias" as if there is one-true-standard against which to compare, you realise the point needs to be made. I think an essential component of any ethical framework chosen to evaluate AI is that admits comparisons with the alternatives. For example, algorithmic decision making compared to the practices (not the theory!) of bureaucracies.

4.4.4. Profiling

Profiling is another name for modelling. What matters is what you *do* with it, as I have argued elsewhere. And it has many advantages[72]. I suggest you find a neutral way to introduce this, and provide a balanced assessment.

---

[66] Administrative Review Council, *Automated Assistance in Administrative Decision Making*, Commonwealth of Australia 2003. See especially page 39, which explains how the issue that such decisions are made by a machine and thus there is nobody to hold to account was dealt with in Australian federal law some 20 years ago by the simple device of holding the secretary of the relevant federal department legally liable: "A decision made by the operation of a computer program under an arrangement made under subsection (1) is taken to be a decision made by the Secretary" (Social Security (Administration) Act, sec 6A; quoted on page 40 of ARC report.

[67] Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Stanford University press, 2010.

[68] *The Control Revolution* op. cit, chapter 8.

[69] Serres refers to it as "mental pollution"; I agree! Michel Serres, *Malfeasance: Appropriation Through Pollution,* Stanford University press, 2010.

[70] The Oxford English Dictionary defines it as "Deceit, fraud, imposture, trickery. 'Something of less value than it seems'; hence, 'something of no value; trifles' (Johnson); worthless stuff, trash, rubbish"; I especially like the verb form: *trump* defined to mean: "to deceive, cheat." Captures it perfectly!

[71] Joseph Kupfer, Privacy, Autonomy, and Self-Concept, *American Philosophical Quarterly,* 24(1), 81-89, 1987; Dominik van Aaken, Andreas Ostermaier and Arnold Picot, Privacy and Freedom: An economic (re_-evaluation of privacy, *Kyklos* 67(2), 133-155 (2014); Jack Hirshleifer, Privacy: Its Origin, Function, and Future, *The Journal of Legal Studies* 9(4), 649-664, 1980.

[72] Quite apart from the obvious fact that it requires you to base your decisions on particular factors, not on individuals. See Frederick Schauer, *Profiles, Probabilities and Stereotypes,* Harvard University press, 2003. A contrary view (which I disagree with) is Bernard E. Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*, University of

5.4 Data integrity, standards and interoperability

This is indeed a crucial problem, and there is precious-little systematic research being conducted on the topic. It is at the core of the Mnemosyne proposal I have already shared.

5.5 Bias – Developing AI (conscious and unconscious bias)

Describing the problem as "bias" is a bad way to do it because it *implicitly* signals that there could be an unbiased way of doing things (there cannot). Talking of "unconscious bias" is even worse given that the whole premise of this rests on scientifically very dodgy ground, especially the notorious implicit association test. Notwithstanding the widespread adoption of this terminology and ideology, it is built upon quicksand.

The latest research shows that even if it is measurable (which it is with poor reliability and repeatability), changing it is possible, but it does not change behaviour[73]. Thus, from any pragmatic perspective, it is merely a dangerous distraction. It is an alarmingly speculative notion on which to base policy decisions. It would be much better to simply say what one is trying to do: make decisions according to a well-articulated and agreed moral framework (which is hardly easy, but at least it does not point one in the wrong direction!)

5.5.1 New forms of discrimination based on aggregation of data

Again I just counsel that you do a fair comparison; all new technologies bring harms. The challenge is to fairly compare them with the alternative. A starting point would be how bureaucracies make decisions. It is deeply ironic that centuries of progress have occurred because of the very deliberate repudiation of the idea that individuals can make decisions non-algorithmically (what else does the Magna Carta offer?). But now there seems to be a desire by some folks to make algorithms make the same mistakes as people!

5.6 Data Governance

I am pleased to see this section heading. It needs to grapple with the *use* of the data, not just that security of it. One can harm people whilst keeping their data absolutely secure.

5.6.1 Trust

Again, refer to the DATA61 science vision, cited earlier.

Chicago Press, 2007. The issue comes down to stereotype accuracy. Stereotypes are *not* bad per se – they are just generalisations. You can have good generalisations or bad ones. See for example Gordon W. Allport, *The Nature of Prejudice,* Addison-Wesley, 1954 (who elegantly defines prejudice as a *faulty* generalisation not corrected by new evidence); Donald T. Campbell, "Stereotypes and the perception of group differences." *American Psychologist* 22(10), 817-829, (1967); Yueh-Ting Lee, Lee J. Jussim and Clark R. McCauley, *Stereotype Accuracy: Toward Appreciating Group Differences,* American Psychological Association, 1995. See also the discussion about the arbitrariness of groups (on which profiling is predicated, and thus on which concerns about profiling is based) in appendix H of Menon and Williamson, *The Cost of Fairness in Binary Classification*, op. cit.

[73] Patrick S. Forscher, Calvin K. Lai, Jordan R. Axt, Charles R. Ebersole, Michelle Herman, Patricia G. Devine5, and Brian A. Nosek, A Meta-Analysis of Change in Implicit Bias*, preprint*, https://osf.io/awz2p/ ; Olivia Goldhill, The world is relying on a flawed psychological test to fight racism, *Quartz,* 3 December 2017, https://qz.com/1144504/the-world-is-relying-on-a-flawed-psychological-test-to-fight-racism/ ; Heather MacDonald, Are we all unconscious racists?, *City Journal*, Autumn 2017 https://www.city-journal.org/html/are-we-all-unconscious-racists-15487.html