Horizon Scanning Series

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

Psychological and Counselling Services

This input paper was prepared by Mike Innes

Suggested Citation

Innes, M (2018). Psychological and counselling Services. 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.

Horizon Scanning Report on AI for Australian Commonwealth Science Council

Submission by J M Innes, Professor of Psychological Science, Australian College of Applied Psychology and Adjunct Professor, University of South Australia

I submit a report related to the performance of people and the interoperability of people and machines in an AI rich world in the context of risks and opportunities within the health sector, specifically the delivery of psychological and counselling services. This sits under section 2.2.3 of the Draft Table of Contents. This submission includes material relevant Section 5.5 on Bias, in particular conscious and unconscious bias in the development and utilisation of AI (risks and benefits).

I am an academic psychologist with long experience in the fields of social and cognitive psychology and in the methodology of social science. I have a long involvement in the training of psychologists at undergraduate and graduate levels. I will comment upon the effects of the methodologies used upon the inferences which are drawn from the research into the effects and consequences of AI.[1.] I include material jointly developed with Dr Ben Morrison of the *Australian College of Applied Psychology*.

Context of discussion on development and consequences of AI: Strong and Weak Models.

- I am not an expert in the technical development of AI systems. In particular, I am not commenting on possible consequences for society or for any practice within society of the *strong* version of AI. That is, I am not able to contribute significantly to the development of any possible AI system which develops consciousness and self-awareness. Such developments will render obsolete all the things that I can contribute upon and will create a scenario akin to that which would follow contact with extra-terrestrial aliens (e.g. Harrison,1997 and many others). I have commented upon the implications of artificial consciousness upon our understanding of humanity, but this is not part of the present brief.
- I shall deal with the possible consequences of developments of the *weak* version of AI, where significant and rapid developments in machine learning and the ubiquity of large computing capabilities and the availability of large data sources have enabled significant developments in what can be programmed into automated services (cf. Clegg, 2017 for a positive view).

Implications for employment, in particular the employment of health professionals.

I address the implications of the development of automation upon employment, in particular the future employment of professionals in the health and helping sectors. These jobs have been portrayed as immune to developments in automation and they are seen as last bastions against the encroachment of automation (cf., Frey & Osborne, 2013; Reese, 2018; Susskind & Susskind, 2015). I argue, however, that:

- Job analysis must take into account the present reality of what those jobs require
- This analysis should not rely upon a superficial and outdated view of the job specifications.

Specifically, psychology is often regarded as a profession with a “calling” (e.g. Seligman, 2018), fulfilling a dream of helping people with a plethora of social skills of extremely high order. The reality in the profession and in the education training establishments does not meet those expectations. An important point to be made is that recent developments in electronic therapy technology are already changing the job characteristics of the psychologist. The technology has already changed the landscape.

The representation of psychology and the helping professions

The job of being a psychologist is being an *expert* in the analysis and understanding of the causes and consequences of human and animal behaviour. Training to be an expert has been traditionally regarded as a process of socialisation into the practices of an expert group; it is a social process of training and service, often with close relationships between the expert trainer and the novices. Crucial to the understanding of expertise is the distinction between *explicit knowledge*, the shared and conscious skills that are necessary for doing the job (written down in text books and portrayed in lectures) and *tacit knowledge*, the deep understanding of the practices acquired through social immersion in the groups who possess it (Collins and Evans, 2007). Becoming a psychologist is not only learning the theories and the methods of the job through explicit tuition. Expertise is based upon immersion with practising psychologists and practicing the skills again and again.

A consequence of the emergence of mass higher education, however, with increasing numbers in universities and private providers, and the difficulty of providing immersion training in skill development, has been the development of lists of skills which are seen as required for performance; *attempting to make the tacit explicit*. These are listed under various rubrics, including “inherent characteristics” and “graduate attributes”. For psychologists and counsellors these include being a good communicator, curious, creative, compassionate, non-judgmental, motivated, able to see patterns and empathic (e.g. Cranney et al., 2009; Thornton 2014). These are also addressed as “soft skills”.

- These characteristics, however, are essentially *personal attributes*, matters which are essential to the *character of the person* and which the person may bring to the job, to the setting in which they are to be trained.
- They can be separated from *skills*, attributes of the job of being a psychologist which can be trained.
- The argument can be put that the skills of a psychologist are essentially based upon the personal attributes of the person who learns the skills; good psychologists are born and not made.
- But the training regimen within the helping professions is to inculcate explicit skills which will enable the person to perform as an expert, without the necessity of acquiring the deep tacit skills.

The job of the psychologist and of the helping professional.

The job specification for a professional psychologist essentially specifies four tasks, whatever may be the area of specialisation (clinical, organisation, forensic, sport etc.). These are

- *Assessment;* the measurement and observation of the client (whether a person or an organisation) to identify the state of the client. This is done by a variety of methods, including systematic behavioural observation, psychometric testing and structured interview.

- *Formulation*; analysis of the data and the development of hypotheses to account for causal relationships between observations and the behavioural, social and economic outcomes that were the primary reason for the client contacting the professional.
- *Intervention*; design of an intervention to change the causal relationships between the measures which are seen as problematic and allow other behaviours and states to occur which will render the problems as less problematic.
- *Evaluation*; measurement of the states after the intervention to ascertain whether change has occurred or not and whether the changes are beneficial or detrimental.

These tasks are central to the training of a professional helper, training based upon evidence derived from the scientific disciplines of psychology, economics (in the case of the organisational psychologist), criminology (for the forensic psychologist), neuroscience, physiology, sociology (in the sense of the development of social systems), and cognitive science. Central to the specification is:

- *The concept of evidence-based practice*, the idea that any practices must be based upon evidence, invariably quantified data from observations and experiments conducted in controlled conditions.
- This enables the range of skills, concepts and practices to be severely curtailed; decisions can be made that particular practices are not sufficiently "evidence-based" and therefore need not be included in the training regimes. An example, within psychology and counselling is the widespread rejection of therapy based upon psychodynamic (Freudian) principles, even though there is copious evidence for their efficacy (e.g. Shedler, 2010; Tracey et al., 2014; Wampold & Imel, 2015; Westen & Bradley, 2005; Woolfolk, 2015). The argument is simply put that some "evidence" is better than others.

The explicit specification of skills enables the automation of skills.

These four characteristics, set out in increasingly specified form with more and more evidence, based upon narrow definitions of what is reliable and valid, leads to the following:

- *Assessment:* Meehl (1954) more than sixty years ago demonstrated that statistical aggregation of assessment (tests or observations) was virtually always superior to aggregation by the clinician. This demonstration has been successively supported (e.g. Dawes, 1994). The development of computer aided tests has increasingly supplanted the provision of assessment by clinicians. Computers can deliver test items and make superior scoring and monitoring of test taking behaviour to anything that a psychologist can do in the room with a client. Item –response analysis technology enables a test to be tailored to the response profile of the client within real-time. The development of virtual reality technology is beginning to extend this even further (e.g. Formosa, Morrison et al., 2018; Parsons & Rizzo, In press). Computer based monitoring, including facial recognition, can be used to assess emotional changes in the client while being assessed, superior to many judgments made by clinicians.
- *Formulation*: the tacit knowledge traditionally regarded as necessary in the development of hypotheses of cause and effect can be seen to be the result of

training in uncontrolled environments wherein there are uncertain relationships between cues and decisions. These uncertain relationships can be identified and the clinician trained to make more and more predictable links (Kahneman & Klein, 2009). Machines can also generalise to previously unseen cases and generate “Probably Almost Correct” (PAC) responses to novel patterns, superior to the human operator. The argument is clear. Given particular assumptions, intuition need not be something mysterious. It can be educated. The work of Tetlock (2005) is often cited to show that experts’ judgments cannot be trusted. This work, however, on the contrary, demonstrates the conditions under which experts are demonstrably able to predict correctly and when not. The automation of intuition can thus be conceived (cf. Morrison et al., 2017).

- *Intervention*: The base of evidence which is used to demonstrate the efficacy of a narrow range of intervention enables the choice of a small number of therapies which can be formulaically treated and clinicians can be trained intensively in those. Specific components of therapy can be identified and introduced at specific points in the therapeutic process which can itself be constrained and presented within limited time frames. The dominance of Cognitive Behaviour Therapy (CBT) is testimony to the prevalence of this methodology and its effects in the clinical profession. The relationship between the therapist and the client (the therapeutical alliance, cf. Wampold & Imel, 2015, for literature) previously regarded as important can be downplayed as relatively less robust than the main effect of the therapeutic technique itself and therefore training in the establishment of this alliance is seen as unnecessary. Akin to the analysis of the British public service being “hollowed out” (Rhodes, 1994), the psychological profession is being hollowed out and rendered replaceable by machine (cf. Innes & Bennett, 2010).

A factor which can be emphasised at this point relates to a central argument used by those who argue that psychologists are not at risk of replacement, namely that psychologists require the attribute of empathy in order to act as psychologists (e.g. Reese, 2018). Without empathy there can be no alliance.

Empathy is described as the ability to “put yourself in their (the client’s) shoes”, being caring, understanding and empowering. It is argued that without this capacity one cannot function as an effective psychologist; it is in fact at the centre of the intensive marketing campaign launched by the *Australian Psychological Society* in 2017.

But there is confusion in the use of the term. It can be used to mean “compassion”, feeling for others and sharing their joy or grief; it is felt *emotion*. It can also mean a sense of cognitive understanding; felt *cognition*. Cognitive understanding can be used to solve problems. The induction of the emotional component on the other hand can lead to bias and misunderstanding (cf. Bloom, 2016). The argument can thus be made that without emotional empathy the trained psychologist is better able to analyse and thereby help a person. And the argument can be developed further. A machine can do this better than a human. Neural nets are already more accurate in detecting facial expressions than humans. AI is already in development to reflect on its own practice and retrain itself to deal with client responses.

- *Evaluation*: The assessment of the benefits or otherwise of the intervention can be addressed in the same manner as the prior assessment. Evaluation can be computer based, linking the pre-measures to post-measures and the data then actuarially examined. This eliminates the biases which have been identified to be present when clinicians make judgments (cf. Lilienthal et al., 2014) and the increasingly dominant literature on the prevalence of unconscious biases in thinking and judgment (e.g. Bargh, 2017; Kahneman, 2011) can be used to argue for the use of explicit and conscious processes for evaluation which can therefore be implemented in automatic cognitive systems. The literature on the other side of the coin is relatively ignored (e.g. Guerin & Innes, 1981; Newell & Shanks, 2014).

Replacement of the psychologists by an automated procedure.

There are six points to be made in the development of this argument.

- The four core elements of a psychologist's job can be specified in sufficient detail to enable an automated version to replace the human being. Not only can the automated version do the job, it will do it better, with less bias, fewer computational and procedural errors in presentation and with no burn out and fatigue. Therefore, there is a clear possibility that psychologists may be replaced by machines. These replacements are in progress, (cf. Boulos et al., 2014: Michie, et al., 2017; Innes & Morrison, 2017 for a review).
- Many psychologists will still be required to continue to develop psychological theory and methodology. Psychologists of particular skill and insight may still be required. But these will be a small proportion of those presently employed in Australia. (Currently there are in excess of 33000 registered psychologists in Australia).
- The reduction in the psychological workforce will be dependent upon the further analysis of the proportion of the psychologist's time is spent with these four components. However, this analysis does not show how extensive is the time spent on these activities. Already psychologists' roles are being changed to monitoring the conduct of an electronic therapeutic intervention rather than act in a face to face role, with equivalent outcomes in the delivery of inter-CBT compared with face to face. *The clock is already ticking*. The replacement level of psychologists will be high, should this analysis turn out to be realised even further in action and adoption of technology.
- There are implications for the education system. Psychology is currently the second largest undergraduate program in Australian universities. While not all students studying psychology wish to become professional psychologists, the large majority of them do so wish. Therefore, the implications for the future training of psychologists are immense (Kennedy & Innes 2005; Innes & Bennett, 2010), not only at postgraduate but at undergraduate levels. There are other views within the discipline which do not predict the wholesale adoption of technology to deliver services. The 2019 Accreditation Standards (*Australian Psychology Accreditation Council*, 2018) are, however, dependent upon the model outlined and adopted in the training of psychologists. Alternative models have no presence in the scenario. The incoming Standards now define the undergraduate program as “pre-professional”, including the elements of the professional activity earlier and earlier in the process and crowding out any possibility of inclusion of broader views of the nature of the discipline.

- The technology is already being adopted and change is occurring internally in the profession. Outside observers are already out of date in their depiction of the nature of the profession of psychology.

There are also other views of the factors which will affect the development of AI in forms which will impact upon the delivery of human services (e.g. Aoun, 2017) but they do not address the fact that the model adopted currently in psychology is based upon the development of a technologically compatible structure which is liable for automation. Levesque (2017) makes an argument for the importance of "common sense" in the development of models, but this can be at least partially opposed by the argument above based upon the possibility of training intuition through the analysis of tacit knowledge. The entire analysis presented here is also based upon the adoption of assumptions about the direction and the causes of the accumulation of scientific knowledge, which are themselves based upon cultural forces and (e.g. Collins & Evans, 2007; O'Gorman, 2017) which can be challenged fundamentally. But the direction within psychology is clear. The current position, upon which this analysis is based, is summarised as:

"Technology can do what therapists cannot, and can do many things better. (cited in Rodham, 2018).

References.

Aoun, J.E. (2017). *Robot-proof: Higher education in the age of artificial intelligence*. Cambridge, MA: MIT Press.

Bargh, J. (2017). *Before you know it: The unconscious reasons we do what we do*. London: Heinemann.

Bless, H., & Burger, A.M. A closer look at social psychologists' silver bullet: Inevitable and evitable side effects of the experimental approach. *Perspectives on Psychological Science*, 11, 296-308.

Bloom, P. (2016). *Against empathy: The case for rational compassion*. New York: Harper Collins.

Boulos, M.N.K., et al. (2014). Mobile phone and health apps. *Online Journal of Public Health Informatics*, 5 (3), e229.

Carr, N. (2010). *The shallows: How the internet is changing the way we think, read and remember*. London: Atlantic Books.

Clegg, B. (2017). *Big data: How the information revolution is transforming our lives*. London: Ikon.

Collins, H., & Evans, R. (2007). *Rethinking expertise*. Chicago: University of Chicago Press.

Cook, T.D., Shadish, W.R., & Wong, V.C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27, 724-750.

Cranney, J., et al. (2009). Graduate attributes of the 4-year Australian undergraduate psychology program. *Australian Psychologist*, 44, 253-262.

Dawes, R. (1994). *House of Cards: Psychology and psychotherapy built on myth*. New York: Free Press.

De Visser. E.J. et al. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology Applied*.

Doyen, s. et al. (2012). Behavioral priming: It's all in the mind, but whose mind? *Plos One*, January 18.

Formosa, N., Morrison, B.W., Hill, G., & Stone, D. (2018). Testing the efficacy of a virtual-reality based simulation in enhancing users' knowledge, attitudes and empathy relating to psychosis. *Australian Journal of Psychology*, 70(1). Doi: 10:1111/ajpy.12167.

Frey, C.B., & Osborne, M.A. (2013). *The future of employment: How susceptible are jobs to computerisation?* Programme on the Impacts of future Technology, University of Oxford.

Guerin, B., & Innes, J.M. (1981). Awareness of cognitive processes: Replications and revisions. *Journal of General Psychology*, 104,173-189.

Harrison, A.A. (1997) *After contact.: The human response to extra-terrestrial life.* New York: Plenum.

Hossenfelder, S. (2018). *Lost in math: How beauty leads physics astray*. New York: Basic Books.

Innes, J.M. (2005). Decline of fact in artefact: Loss of control in social psychological studies. *Australian Journal of Psychology*, Supplement, 57, 89.

Innes, J.M., & Bennett, R.G. (2010). Training the professional psychologist: Is there a need for a reappraisal of the nature of education within a scientific paradigm? *Fourth International Conference on Psychology Education*, Sydney.

Innes, J.M., & Morrison, B. W. (2017). Projecting the future impact of advanced technologies on the profession: Will a robot take my job? *InPsych*, Australian Psychological Society, 39 (2), April, Pp. 34-35.

Kahneman, D. (2011). *Thinking fast and slow*. New York. Farrar, Straus and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise.: A failure to disagree. *American Psychologist*, 64, 515-526.

Kennedy,B.,& Innes, M. (2005). The teaching of psychology in the contemporary university: Beyond the accreditation guidelines. *Australian Psychologist*, 40, 159-169.

Klein, O., et al. (2012). Low hopes, high expectations: Expectancy effects in the replicability of behavioral experiments. *Perspectives on Psychological Science*, 7, 572-584.

Leigh, A. (2018). *Randomistas: How radical researchers changed our world*. Melbourne: University of Melbourne Press.

Levesque, H.J. (2017). *Commonsense, the Turing test and the quest for real AI*. Cambridge, MA: MIT Press.

Lilienfeld, S.O., Ritschel,.A., Lynn, S.J., Cautin, R.L., & Latzman, R.D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, 9 (4) 355-387.

Meehl, P.E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Michie, S., et al. (2017). Developing and evaluating digital interventions to promote behavior change in health and healthcare. *Journal of Internet Research*, 19 (6), e 232. Doi: 10.2196/jmir.7126.

Morrison, B.W., Innes, J.M., & Morrison, N.M.V. Current advances in robotic decision-making: Is there such a thing as an intuitive robot? Australian Psychological Society Industrial and Organisational Psychology Conference, Sydney, July 2017.

Newell, B.R., & Shanks, D.R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, 37, 1-61.

O'Gorman, F. (2017). *Forgetfulness: Making the modern culture of amnesia*. London: Bloomsbury.

Parsons, T.D., & Rizzo, A.A. (In press). A virtual classroom for ecologically valid assessment of attention deficit/hyperactivity disorder. *Virtual reality technologies for health and clinical applications: Psychological and neurocognitive interventions.*

Reese, B. (2018). *The Fourth Age*. New York: Atria.

Rhodes, R.A.W. (1994). The hollowing out of the state: The changing nature of the public service in Britain. *The Political Quarterly*, 65, 138-151.

Rodham, K. (2018). Self-management. *The Psychologist*, July. Pp.34-37.

Rosenthal, R., & Rosnow, R. (1969). *Artefact in behavioural research*. New York: Academic Press.

Seligman, M. (2018). *The hope circuit: A psychologist's journey from helplessness to optimism.* Sydney: Penguin Life.

Shadish, W.R., & Cook, T.D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, 60, 607-629.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologis*t, 65, 98-109.

Susskind, R., & Susskind, D. (2015). *The future of the professions: how technology will transform the work of human experts.* Oxford: Oxford University Press.

Tetlock, P.E. (2005). *Expert political judgment*. Princeton: Princeton University Press.

Thornton, A. (2014). Employment, training and professional development of provisional psychologists in the corporate sector. *In Psych*, Australian Psychological Society, December.

Tracey,T.J.G, Wampold, B.E., Lichtenberg, J.W., & Goodyear, R.K. (2014). Expertise in psychotherapy: An elusive good? *American Psychologist*, 69 (3), 218-229.

Wampold, B.E., & Imel, Z. E. (2015). *The great psychotherapy debate*. New York: Routledge.

Westen, D., & Bradley, R. (2005). Empirically supported complexity: Re-thinking evidence based practice in psychotherapy*. Current Directions in Psychological Science,* 14, 266-271.

Woolfolk, R.L. (2015). *The value of psychotherapy*. New York: Guilford.

Methodological note.

Critical attention needs to be paid to the nature of the evidence used to support arguments in this area. The first point is that there is a strong attraction to the use of the randomised experiment in studies used to support various features of human/robot interactions, which claim either to enhance or reduce the expectations that people have about robots and the consequences of such interactions. The preference for the randomised control trial (e.g. Leigh, 2018) does not address the fact that there is clear evidence of the presence of strong threats to the validity of the data from experiments carried out in non-laboratory conditions. The work of Campbell and his co-workers (Cook et al., 2008; Shadish & Cook, 2009; Shadish et al., 2002) among many others, shows that there are systematic biases in supposedly randomised experiments which result in severe doubts about the validity of the data.

More fundamentally, the use of laboratory experiments to claim certain benefits of interventions in the interactions with robots (e.g. de Visser, et al. 2016) pays no attention to the body of literature created in experimental social psychology that demonstrated the biases that social interaction between experimenter and respondent can severely affect the outcomes of experiments and prevent any clear inference of validity (e.g. Innes, 2005; Rosenthal & Rosnow, 1969). The irony is that even in social psychology there has been a failure to remember the past and learn from the mistakes so that much of the literature in this field is now tainted (cf. Klein et al., 2012) and the so-called “silver bullet” of the experimental method in social psychology failed to have an effect (cf. Bless & Burger, 2016). The double irony, which is of significance in this field of interaction with AI, is that people are now thinking about the importance of biases affecting the ways in which research is developed (cf. Hossenfelder, 2018) and refer to work which itself is biased because of the failure to implement appropriate experimental controls. Work on bias and the effects of unconscious motivation affecting judgment which is suggested to have relevance to studies of human versus computer judgment referred to in this submission is one area where significant doubts about validity have been raised. (e.g. Doyen et al., 2012). The detail, wherein lies the devil, is always in the method section of the journal article and not in the abstract.