PSYCHOLOGY MISCELLANY

No.177 - 1st January 2023

Kevin Brewer

ISSN: 1754-2200

orsettpsychologicalservices@phonecoop.coop

This document is produced under two principles:

1. All work is sourced to the original authors. The images are all available in the public domain (most from http://commons.wikimedia.org/wiki/Main_Page). You are free to use this document, but, please, quote the source (Kevin Brewer 2023) and do not claim it as you own work.

2. Details of the author are included so that the level of expertise of the writer can be assessed. This compares to documents which are not named and it is not possible to tell if the writer has any knowledge about their subject.

Kevin Brewer BSocSc, MSc

An independent academic psychologist, based in England, who has written extensively on different areas of psychology with an emphasis on the critical stance towards traditional ideas.

A complete listing of his writings at http://psychologywritings.synthasite.com/. See also material at https://archive.org/details/orsett-psych.

CONTENTS

Page Number

1. AUTHENTICITY AND IDENTITY TODAY

1.1. AUTHENTICITY

Authenticity "typically connotes a coveted status, a 'positive value, revered and sought without ambivalence' (Lindholm 2013...)" (Schwarz and Williams 2021 p1). It is mentioned in everyday life in many and varied situations, including "a sought-after characteristic and marketing claim in the culinary sphere - imbued into local foods, craft beverages, traditional recipes, and 'ethnic' cuisines... Travellers seek to locate authenticity when they visit heritage sites, regional festivals, and theme parks... Whether or not someone is an authentic member of some social group or sub-culture often evokes spirited debate, sometimes even outright conflict... Authentic self-expression is a concern within the confessional cultures associated with social media, memoir, and reality television..." (Schwarz and Williams 2021 1) [1].

Authenticity is synonymous, according to Lindholm (2013), with "sincere, true, honest, absolute, basic, essential, genuine, ideal, natural, original, perfect, pure, real, and right" (quoted in Schwarz and Williams 2021). But how to define it as a concept? Martin (2014) felt that it was "so bloated with meanings that the hope of true definition is nearly futile" (quoted in Schwarz and Williams 2021).

Schwarz and Williams (2021) concentrated on authenticity in relation to identity. But, even here, there is diversity in its use. For example, Pierce (2015)

[1] Pine and Gilmore (2013) explained authenticity in relation to consumerism: "In a world increasingly filled with deliberately and sensationally staged paid-for experiences, people increasingly see the world in terms of real and fake, and want to buy the real from the genuine, not the fake from some phony. They now decide where and when to spend their money as much if not more than they deliberate on what and how to buy. But in a world of experiences – an increasingly unreal world – consumers choose to buy or not buy based on how real they perceive an offering to be. In other words, authenticity has become the new consumer sensibility" (p29).

But these authors asked, "what exactly is authenticity, in business terms?". They answered: "It's purchasing on the basis of conformance to self-image. Economic offerings that correspond in both depiction and perception to one's self-image are perceived as authentic. Those that do not match to a sufficient enough degree to generate a 'sympathetic vibration' between the offering and the buyer are viewed as inauthentic. So consumers now purchase offerings based on how well they conform to their own self- image, both who they are and who they aspire to be – with lightning-quick judgments of 'real' or 'fake' hanging in the balance" (Pine and Gilmore 2013 p29).

distinguished "first-person" authenticity ("subjective") from "second-person" authenticity ("inter-subjective"), while Carroll (2015) talked of "moral" and "type" authenticities (Schwarz and Williams 2021).

In relation to identity, authenticity, simplistically, is associated with the "true self" (ie: "the characteristics, roles, or attributes that define who you really are – even if those characteristics are different than how you sometimes act in your daily life"; Seto and Schlegel 2018 quoted in Schwarz and Williams 2021). In everyday language, "being true to yourself".

Popular self-improvement literature is "flush with the idea that people can become authentic" (Schwarz and Williams 2021 p6). Schwarz and Williams (2021) noted: "The idea that there exists an authentic core to one's being (often hidden or untapped) may inspire and encourage, but may also propel individuals 'on an endless journey of selfhood, in which authenticity is always just out of reach, like the mythic apples of Tantalus' (Fordahl 2018...)" (p6).

More widely, "authentic identity" can be a way for individuals to demarcate themselves (and their group) from others. "For example, Greenebaum (2012) shows how 'ethical vegans' differentiated themselves from 'health vegans' and 'plant-based eaters' who fall short of the moral criteria they put in place to establish themselves as authentic. Yet when some of those 'ethical vegans' wore leather shoes, ate honey, or used pharmaceuticals, they afforded their transgressions some leeway, as reasonable concessions in a world where 'pure' veganism is practically impossible" (Schwarz and Williams 2021 p7).

Such a view is grounded in essentialism - ie: "the idea of authenticity gains its force from essentialism, for the possibility of a 'real' or 'genuine' group member relies on the belief that... [they] possess inherent and perhaps even inalienable characteristics criterial of membership" (Bucholtz 2003 quoted in Schwarz and Williams 2021).

1.2. TRIANGULAR THEORY OF SELF

Wang (2022) proposed "a triangular theory of self to characterise the sense of selfhood unique in the era of social media" (p1). The three components are:

i) The represented self – This is "the characteristics, roles, and experiences of oneself as perceived and encoded by the person who acts as an agentic experiencer and knower" (Wang 2022 p2). This is similar to the traditional idea of the subjective self (ie: autobiographical and personal experiences "'located' in the private mind and brain of the person"; Wang 2022 p3).

At the same time, individuals share elements of this in blogs and personal postings on social media. Thus, "the represented self is receptive to the context of social media and the virtual community. Although private to the individual mind, it is externalised in the social media era, a motivated process that, in turn, shapes self-representations. The person plays an active role in creating the digital extension of the self that is registered online" (Wang 2022 p5).

ii) The registered self – This is the aspect of the self shared on social media. It is "public by default and interactive in nature... [and] may bear some resemblance with what is traditionally portrayed in memoirs or autobiographies, whereby, in both cases, the person tells his or her stories to an intended audience" (Wang 2022 p5).

iii) The inferred self - This is the self "as viewed and interpreted by the virtual audience" (Wang 2022 p2). So, "[W]hile the person is an active agent in selectively sharing autobiographical information online, the receiver or consumer of the information, namely the virtual audience, also play an active role in inferring and working out who the person is" (Wang 2022 p7).

This means, though, that the person "may anticipate the expectations of the audience when making future posts about themselves and further engage in self-censorship and strategic self-presentation. For instance, individuals post more frequently when they feel excited about the Likes and comments they receive, while they change the content of their future posts when they feel sad for the lack of audience response" (Wang 2022 p8).

Online sharing also aids memory. For example, Wang et al (2017) asked participants to keep a private diary for one week, and to note what information within it they shared online. In a later memory test, individuals recalled more of the information posted online than not from the diary.

The three components of the self interact to produce the sense of self that Wang (2022) saw as unique to the social media era. Put simply, the self is experienced differently through the all-pervasive interaction with social media.

I would say that Wang (2022) has not paid enough attention to aspects of what or how information should be shared. For example, the predominance of "happy" [2] and "successful" postings suggesting a "great life". The pressure to conform in this and other ways in consumerism has been called "conforming individualism" [3], or, my term, "conforming unique".

I would also use the term "appropriate conformity" to describe, for instance, the posting of pictures to get "Likes" and to test what is an "acceptable identity". Though there are spaces for individuals to share negative and "minority" interests, the common use of phrases like "be yourself" and "be true to yourself" do not actually mean that. Individuals who are "too" different in being true to themselves are often heavily criticised and marginalised.

1.3. PHOTOGRAPHS ON SOCIAL MEDIA

Sharing personal photographs on social media is ubiquitous, but how does this behaviour impact an individual's recollection of their past?

Fawns (2020) outlined four distinct components of the act of photography - capturing (ie: taking the photograph), organising, viewing, and sharing. Stone and Zwolinski (2022) added an extra component of manipulating in between capturing and organising.

Stone and Zwolinski (2022) discussed each component in relation to memory:

a) Capturing - Henkel (2014), for instance, asked participants to photograph some objects, and not others, in a museum. Recognition memory for the photographed objects was poorer, which suggested that "the photograph is acting as an externalised memory device" (Stone and Zwolinski 2022 p4). But this study found that recall was better for photographed objects when the camera had been zoomed-in. So, Stone and Zwolinski (2022) proposed, "it is the cognitive process (or lack thereof) individuals undertake during the act of taking photographs that

[2] I would use the term "advert happy" to describe the expectations of joy and happiness to mirror that shown by women in particular in advertisements.

[3] This is "the tendency for individualism to set up a certain model for everybody to follow so that one may become an individual like everybody else" (Montuori and Purser 2000).

ultimately shapes one's memory for the captured moment" (p4).

b) Manipulating - Digital photographs can be adjusted to make them more interesting/attractive or remove unwanted elements (or people). So, "the fact that individuals can and are manipulating their own photographs may figure heavily in how such pictures come to shape the way they remember their personal past (eg: false recollections...) as well as influence their well-being at the moment of revising the manipulated picture (eg: 'I was so thin back then!'...)" (Stone and Zwolinski 2022 p5).

c) Organising - The organisation of photographs online allows greater variety than the traditional use of photo albums, say, but if "an individual's idiosyncratic organisation is disrupted, individuals tend to recall less information" (Stone and Zwolinski 2022 p5).

d) Viewing - Viewing photographs can improve memory for them, but manipulated ones can implant false memories (Stone and Zwolinski 2022).

e) Sharing - The research is mixed around sharing personal photographs and memory. Some studies find improved recall for shared over unshared information/photographs, some studies the opposite. "However, there is evidence suggesting that sharing photographs via social media may facilitate the recall of the shared events" (Stone and Zwolinski 2022 p6).

It is interesting to see what is not shared - ie: photographs that "individuals remain 'social media silent' about" (Stone and Zwolinski 2022 p8).

1.4. REFERENCES

Bucholtz, M (2003) Sociolinguistics nostalgia and authenticity of identity Journal of Sociolinguistics 7, 3, 398-416

Carroll, G.R (2015) Authenticity: Attribution, value, and meaning Emerging Trends in the Social and Behavioural Sciences (https://onlinelibrary.wiley.com/doi/book/10.1002/9781118900772)

Fawns, T (2020) Blended memory: A framework for understanding distributed autobiographical remembering with photography Memory Studies 13, 6, 901-916

Fordahl, C (2018) Authenticity: The sociological dimensions of a politically consequential concept The American Sociologist 49, 2, 299-311

Greenebaum, J (2012) Veganism, identity and the quest for authenticity Food, Culture and Society 15, 1, 129-144

Henkel, L.A (2014) point-and-shoot memories: The influence of taking photographs on memory for a museum tour Psychological Science 25, 396-402

Lindholm, C (2013) The rise of expressive authenticity Anthropological Quarterly 86, 2, 361-395

Martin, M.S (2014) Authenticity, tourism, and Cajun cuisine in Lafayette, Louisiana. In Cobb, R (ed) The Paradox of Authenticity in a Globalised World New York: Palgrave Macmillan

Montuori, A & Purser, R (2000) In search of creativity: Beyond individualism and collectivism Paper presented at Western Academy of Management Conference, Hawaii

Pierce, A.J (2015) Authentic identities Social Theory and Practice 41, 3, 435-457

Pine, B.J & Gilmore, J.H (2013) The experience economy: Past, present and future. In Sundbo, J & Sorensen, F (eds) Handbook on the Experience Economy Cheltenham: Edward Elgar Publishing

Schwarz, K.C & Williams, J.P (2021) Introduction to the social construction of identity and authenticity. In Williams, J.F & Schwarz, K.C (eds) Studies in Social Construction of Identity and Authenticity Abingdon: Routledge

Seto, E & Schlegel, R.J (2018) Becoming your true self: Perceptions of authenticity across the lifespan Self and Identity 17, 3, 310-326

Stone, C.B & Zwolinski, A (2022) The mnemonic consequences associated with sharing personal photographs on social media Memory, Mind and Media 1, e12, 1-11

Wang, Q (2022) The triangular self in the social media era Memory, Mind and Media 1, e4, 1-12

Wang, Q et al (2017) Externalising the autobiographical self: Sharing personal memories online facilitated memory retention Memory 25, 6, 772-776

2. ARTIFICIAL INTELLIGENCE: MAKING SENSE OF ITS USE IN HEALTH AND SOCIETY

2.1. INTRODUCTION

"There is a multiplicity of definitions of AI [artificial intelligence], as an umbrella term for machine learning, autonomous systems, intelligent data mining and smart information systems... AI is a catch-all term for diverse sets of techniques as well as research agendas, and thus for a large number of sub-fields such as: cognitive computing (algorithms that reasons and understand at a higher (more human) level), machine learning (algorithms that can teach themselves tasks), augmented intelligence (cooperation between human and machine) and AI robotics (AI embedded in robots)" (Ethics and Society et al 2021 p12) (appendix 2A).

The following areas are particularly relevant to health:

a) Machine learning to spot patterns in data, and machine reasoning to make optimal decisions.

"Unsupervised learning algorithms, which aim to discover inherent structures in data without using pre-existing categories, are notoriously inscrutable even to their designers... In this context, the idea of using such unsupervised machine learning to discover 'brain signatures' of neurological conditions or 'biomarkers' of mental disorders appears problematic" (Ethics and Society et al 2021 p21).

For example, in mid-2021, data were released showing how AI was used to predict over 20 000 human proteins (Tunyasuvunakool et al 2021). A machine-learning tool called "Alphafold" was developed by a company called "DeepMind", which was trained on DNA sequences. Predicting the 3D shape of protein folds "by hand" has been labour intensive in the last half of a century (Editorial 2021b).

Computational biologist, Christine Orengo, called it "totally transformative" to have so much data, while another computational biologist, Mohammed AlQuraishi talked of a "paradigm shift" (Callaway 2021). But the AI "doesn't (at least for now) explain how, or why, these results happened" (Editorial 2021b p625) [4].

b) Brain modelling and simulation.

There are ethical issues in relation, for example, to "the conclusion it may allow drawing with regards to consciousness" (Ethics and Society et al 2021 p22).

c) Brain-computer interface (BCI) or "neurorobotics".

Everyday or "mundane" forms of AI and robotics are expected to increase in the near future. "These can include (among others), applications that seek to optimise workflows, services or information systems (in households, firms, health care institutions, traffic systems, government units, etcetera). The fact that these technologies promise new solutions, increased efficiency and lower costs, does not mean they are inevitably beneficial or without risks and unintended disruptive effects" (Ethics and Society et al 2021 p24).

Crawford (2021a) added to machine learning, the social practices involved in designing systems, and the process of "massive data harvesting".

She argued that "far from being something abstract

[4] A key problem for machine learning is "catastrophic forgetting". "This is the tendency of artificial networks to forget previously learned information when they learn something new. Using more complex artificial neurons seems to get around this by allowing different regions of the network to specialise at different tasks" (Gent 2022 p40).

and objective, [AI] is both material and intrinsically linked to power structures. The way it is made involves extracting resources from people and the planet, and the way it is used reflects the beliefs and biases of those who wield it. Only when we come to terms with this... will we able to chart a just and sustainable future with AI" (p47).

Thinking of AI as like human intelligence is a problem, Crawford (2021a) argued, because of "enchanted determinism": "the belief that...[AI] systems are both magical and at the same time can provide insights about all of us in ways that are superhuman. This means that we're not expecting these systems to produce forms of bias and discrimination. Nor do we focus on the ways in which they're constructed and their limitations" (p48).

Artificial General Intelligence (AGI) is AI that "go well beyond being good at one specific task, but can instead do anything a human can" (The leader 2021 p5). It is hoped that scaling up neural networks will make them more "brain-like" (Rorvig 2021).

Erin Young agreed: "Technology is not neutral. It's shaped by the people that build the technologies, shaped by their choices and their values... There is mounting evidence that suggests the under-representation of women in AI roles within tech companies results in feedback loops. AI systems are not objective. So when bias goes in, bias comes out" (quoted in Titcomb 2021). This comment came after Facebook was accused on using algorithms that confirmed gender bias. Algorithms that "optimise for ad delivery" had led to job advertisements for mechanics and pilots being seen almost exclusively by men, and advertisements for nursery nurses by mostly women. There is a call to "de-bias" algorithms (Titcomb 2021).

AI software that assesses emotions from facial features is growing in popularity, but there is concern over its use and misuse (Crawford 2021b). In fact, a review by Barrett et al (2019) questioned the reliability of emotion-recognition from facial expressions generally.

2.2. TRUST AND TRANSPARENCY

"Trust and trustworthiness are central to discussions on the ethics of AI" (p11), pointed out Ethics and Society et al (2021). One way to understand trust in AI is the concept of "epistemic trust" (Koenig and Harris 2007) - ie: "people’s willingness to accept that the knowledge and information provided by, for example, scientific or diagnostic devices, social media or intelligent systems such as AI is accurate and reliable and can be used as a basis for learning and decision-making" (Ethics and Society et al 2021 p14).

Ethics and Society et al (2021) stated: "Another central question is who will keep the oversight on how, by whom and for which purposes digital data are collected and processed? Current systems of data collection and use are primarily designed as 'one-way mirrors'. Citizens and consumers can always be identified, but they never know which kind of information is gathered, how these data are used and for which purposes, and which types of organisations, companies or governments units are involved in these processes. Such practices are fundamentally dangerous, because they compromise anonymity, informed consent and security. It also creates distrust and suspicion" (p16).

Transparency goes with trust, Ethics and Society et al (2021) emphasised. In relation to AI, it conceptualised in different ways: "Sometimes the term refers to a lack of deception. This implies that the internal workings of AI technology should be open to inspection and evaluation. At other times, the concept refers to a mechanism to report reliability, ie: the provision of information on the system's tendency to produce errors. At still other times, it has been conceived as a means to communicate unexpected behaviour, to account for the conditions and risks when AI acts differently than expected. Most frequently, however, transparency is used to refer to the need to make decision-making processes accessible to users, so that they can understand and judge how an autonomous system has reached a certain decision" (Ethics and Society et al 2021 p17).

Spiegelhalter (2020) offered seven questions to ask about the trustworthiness of algorithms:

"1. Is a new algorithm any good when tried in new parts of the real world?

2. Would something simpler and more transparent and robust, be just as good?

3. Could I explain how it works (in general) to anyone who is interested?

4. Could I explain to an individual how it reached

its conclusion in their particular case?

5. Does it know when it is on shaky ground, and can acknowledge uncertainty?

6. Do people use it appropriately, with the right level of scepticism?

7. Does it actually help in practice?" (Ethics and Society et al 2021 p18).

More widely, the "lack of transparency enables possibilities for abuse and manipulation, for example misinformation in elections or the strategic manipulation of public opinions which can lead to a loss of empathy or feelings of alienation. Abuse and misuse as a major concern and uncertainty" (Ethics and Society et al 2021 p19). This adds a need for accountability.

Slonim et al (2021) reported an experiment where a machine powered by AI ("Project Debater") debated live with a human. The machine could not always match the coherence and fluency of the human speaker (Editorial 2021a). There could be a time when AI is as persuasive as humans, and so transparency is needed so that individuals know whether they are interacting with an AI or a human (Editorial 2021a).

AI with elements of theory of mind, which could be explicitly manipulative, are being developed (eg: SEAI; Cominelli et al 2018) (Editorial 2021a).

2.2.1. Bias

A neural network developed with passport-size images of faces from the Internet (Hundt et al 2022) given a choice of two images was more likely to associate a Black man with "criminal" than a White man, and a Black or Latino woman with "homemaker", but a White woman less so with "doctor" than a White man (Stokel-Walker 2022).

Bailey et al (2022) analysed 296 billion web pages to see how gender neutral words like "people" and "humanity" were embedded in text. The frequency of occurrence of these words better matched the context of words like "men" and "male" than "female" and "women".

This is a reflection of sexist bias in society, but, worryingly, "the same texts are used to train AI tools that will inherit this bias, including translation websites and conversation bots" (Sparkes 2022a p23).

Carissa Veliz (at the University of Oxford) argued that AI training data should be created from scratch (rather than using existing data from the Internet) "so that they don't include harmful biases, but [she] concedes this is a daunting task" (Sparkes 2022c).

2.3. "MIND READING" ALGORITHMS

Electro-encephalography (EEG) measures electrical activity generated by groups of neurons in the brain. It can be used to find the pattern of brain activity in response to particular audio stimuli (ie: music), for example.

It has been found that the brain anticipates melodic information as well as responding differently to tempo and beat frequency variations (Sonawane et al 2021). "However, work done on pattern of brain activity reflecting neural entrainment to music listening and its recognition is still at an early stage. These patterns are very much intricated and thus it is hard to interpret what is happening in the human brain when a person is listening to a song. Moreover, aesthetic experience associated with music listening is highly subjective - ie: it varies from person to person and also from time to time depending on various contextual factors such as mood of the individual who is listening to music" (Sonawane et al 2021 pp154-155).

Sonawane et al (2021) set out to find identifiable EEG patterns to music stimuli, and differences between individuals. Twenty participants were played short extracts (120 seconds) of twelve popular music stimuli (eg: dance; rock; hip hop) while EEG readings were taken from 128 points on the skull, and afterwards the songs were rated for enjoyment and familiarity. The music pieces varied in tone, rhythm and pitch, and some of them included vocals. Machine learning was then used to analyse the EEG patterns.

The algorithm was able to distinguish the particular piece of music being listened to from the EEG patterns better than chance. However, the accuracy was better for some pieces of music than others, and between individuals. Overall, the key finding was that identifiable patterns of EEG activity were generated by the brain when listening to different types of music, and these patterns vary from individual to individual.

The optimal accuracy of music identification from EEG patterns was 85%, which was better than two previous studies using similar machine learning models:

i) Yu et al (2018) - Ten-second extracts of eight different categories of sound (eg: chant; yodelling; rapping) with nine male participants; accuracy 69%.

ii) Stober et al (2014) - 24 music stimuli played at different pitches; accuracy 24% (Sonawane et al 2021).

Sonawane et al (2021) ended with the slightly disturbing point that the "results achieved in this paper are highly encouraging and provides an essential step towards an ambitious mind-reading goal" (p160).

2.3.1. "Beauty Centre"

Is there an area of the brain linked to perception of beauty. Recording brain activities while individuals look at beautiful things or people is the main method of study. The findings, however, are not in agreement. Chuan-Peng et al (2020) explained: "There is still no consensus about whether there is a unique brain representation specific to beauty... Inconsistent findings across studies might suggest that there are beautify form-specific modules in the brain. The discrepant results also might reflect variations of tasks and relevant parameters used in these studies, although there might be a common beauty-representation in the brain" (p1200).

in simple terms, a single brain region (eg: medial orbito-frontal cortex; mOFC), or a network of connected regions (eg: "the interactions between sensory-motor, emotion-valuation, and meaning-knowledges neural systems; Chuan-Peng et al 2020 p1201).

Chuan-Peng et al (2020) proposed meta-analysis as the solution, and they collected 49 neuroimaging studies (published up to mid-December 2020) to synthesise. Twenty articles on visual arts and beauty, and 29 with human faces. The focus was on "studies that compared the neural activities associated with the faces and visual arts (ie: paintings, visual patterns, architectures, and dances) rated as beauty to those rated as non-beauty" (Chuan-Peng et al 2020 p1201) (table 2.1).

The scan data were synthesised using an activation likelihood estimation algorithm, which was "designed to identify areas that exhibit a convergence of reported co-ordinates across experiments that is higher than expected under a random spatial association" (Chuan-Peng et al

* Only studies involving whole-brain analyses.
* Use of standard neuroimaging procedures.
* Non-expert adults aged 18-50 years old (ie: not art experts).
* The experience of beauty in response to stimuli rated as beautiful versus not beautiful.

Table 2.1 - Key inclusion and exclusion criteria for studies in Chuan-Peng et al's (2020) meta-analysis.

2020 p1203).

No common brain regions were found for both beauty in the visual arts and in faces, overall. But the anterior medial prefrontal cortex (aMPFC) was consistently activated by the beauty of visual arts, while two areas were activated by the beauty of faces (the ventro-medial prefrontal cortex; vMPFC, and the left ventral striatum).

Any meta-analysis is only as good as the studies involved, and there was variety between them in methodology (eg: small sample sizes; differences in data analysis and neuroimaging software implementation). Studies not reporting any brain activation were not included in the meta-analysis, which is a potential selection bias, and that such studies are probably not published anyway is a publication bias issue (Chuan-Peng et al 2020).

Chuan-Peng et al (2020) ended: "Our meta-analytic results revealed distinct neural specificities for visual art and face beauty but lacked evidence for the common neural basis of visual beauty. This null result suggests that the available data did not support the notion of the existence of a common brain for processing different forms of beauty. To support such a common neural basis for beauty, more rigorous studies are needed" (p1211).

2.3.2. Research Ethics

Samuel et al (2021) noted: "With the advent of the 'digital turn', traditional research methods have been reimagined, with new forms of data available to study (social media, blogs, data from wearable devices, and electronic health records) and new methodological tools to help researchers to access, process, and harness this data (artificial intelligence [AI], data modelling). This has presented specific ethical challenges, and in some instances, has disrupted 'traditional' understandings of research ethics, problematising notions of consent and privacy, and raising questions around what constitutes human participant research (we prefer the notion 'data subjects'..." (p326).

These authors focused on AI, with particular reference to population health research, by interviewing eighteen researchers in the field. A number of themes related to research ethics emerged, including:

i) "Data governance" (eg: data subjects' privacy) - For example, "Interviewee 4" said: "it's quite easy to identify people even if the data has been de-identified so I am not able to add anything to that data. Somebody has to look at everything... if I want to load up a set of codes to run against that data, somebody checks that... I can't share it [the data] with anybody who doesn't have permission... if I want to generate any reports... again, there is a secure control process... to ensure privacy and confidentiality" (p328).

ii) Research ethics committees (RECs) - The experience of the interviewees with RECs was "around data access and not about the software very much" ("Interviewee 12"; p5). While "Interviewee 17" noted that "people are intimidated by it [AI} and they are just 'well, this is the computer stuff and we will focus on the things we understand'" (pp329-330).

iii) Societal use of AI systems - "Interviewee 8" described the problem: "what we see emerging at the moment is delegating more authority for making decisions [in healthcare] to what might be termed AI... how do we know that machines are safely making decisions. How do we know that these machines work, because they're generally quite different to a human judgement" (p330).

"Interviewee 13" described a difference between medications and AI: "[a drug is only] licensed for a particular indication and a particular set of populations, [similarly AI systems are only developed on specific datasets], and once it's in the market it can be used by clinicians for any indication and any population [...] so bad things happen" (p331).

Overall, Samuel et al (2021) found a "separation between the ethical practice of research and research use [which] is not necessarily problematic in and of itself. However, the differentiation did permit another issue to be exposed. That is, while the governance of research ethics is tightly regulated, the ethics governance of research use is not. This is problematic if researchers are being called upon to consider the ethics of the societal impact of the research, and take responsibility for this, because... it leaves them with little way to enact such responsibilities" (p333).

2.3.3. Ethics of Big Data

Ethics around research with humans have concentrated on the potential harms to participants (eg: physical or psychological pain). But today "big data" research projects, particularly biomedical, "leverage unconventional data sources (eg: social media), partially inscrutable data analytics tools (eg: machine learning), and unprecedented volumes of data. Moreover, the evolution of research practices and new methodologies such as post-hoc data mining have blurred the concept of 'human subject' and elicited a shift towards the concept of data subject — as attested in data protection regulations" (Ferretti et al 2021 pp1-2).

Researchers can now access data on a person from databases without ever meeting that person. "As such, the nature of risk involved in this new form of research changes too. In particular, it moves from the risk of physical or psychological harm towards the risk of informational harm, such as privacy breaches or algorithmic discrimination. This is the case, for instance, with projects using data collected through web search engines, mobile and smart devices, entertainment websites, and social media platforms" (Ferretti et al 2021 p2).

Ferretti et al (2021) considered the weaknesses of the traditional Ethics Review Committee (ERC) in this context. Prior to beginning a project, researchers provided the relevant ERC with an outline of ethical issues involved and how they would deal with them, and an independent assessment was made.

Ferretti et al (2021) described two categories of weaknesses with the ERC in the case of "big data":

i) Persistent weaknesses - Weaknesses in the current system that are exacerbated by big data research (eg: the overwhelming number of projects to review and the time required; inconsistent assessment criteria between different ERCs).

ii) Novel weaknesses - New problems created by big data projects.

a) Purview weaknesses - Some big data projects may not see it as necessary to apply to an ERC (eg: use of anonymous data). Ferretti et al (2021) pointed out that "using anonymised data should not be deemed oversight-free by default, as it is increasingly hard to anonymise data. Technological advancements might soon make it possible to re-identify individuals from aggregate data sets" (p6).

b) Functional weaknesses - The ERC as inadequate to assess big data projects (eg: lack of expertise in big data). For instance, "the distinct methodology of big data studies (based on data aggregation and mining) requires a specialised technical expertise (eg: information systems, self-learning algorithms, and anonymisation protocols)" (p7).

Ferretti et al (2021) made some recommendations including "new ethical guidelines and new ethical assessment tools to safeguard society from novel risks brought by big data research" (p10), the inclusion of appropriate expertise on ERCs, and the involvement of non-specialist in the process to increase transparency (ie: ordinary ("data subjects") people).

2.3.4. Metaphor

Cobb (2020) outlined a history of the metaphors used to understand the human brain (eg: machines; telecommunication networks). "Metaphors change how science is done, by licensing new interpretations or inspiring new experiments" (Casper 2020 p23).

But also "metaphors conceal as much as they reveal. The ideas that they so persuasively represent often ignore key elements. Comparing the brain to a computer is beguiling, but neglects that brains are also organs, and aware ones at that" (Casper 2020 p23).

Casper (2020) continued: "researchers should acknowledge that although certain word choices seem innocent, many carry malign overtones. Ideas of the brain have often embedded inequities and prejudices about race, class, gender, sexuality and agency" (p24) (eg: "primitive").

2.3.5. Animals

AI programmes that recognise facial expressions are being developed for humans, and also for non-humans. But expressions my not mean the same thing. For example, the bearing of teeth in humans is interpreted as smiling, but in non-human primates it is more commonly associated with aggression (Neethirajan 2021).

If nothing else, the ability to identify stress in captive and domesticated animals would help in their welfare. With this in mind, Neethirajan (2021) described the WUR Wolf (Wageningen University and Research: Wolf Mascot) - a digital facial expression system for farm animals.

The ability to recognise thirteen facial actions and nine emotional states in cows and pigs was based on parameters like ear postures, and eye white regions (table 2.2).

Cow

- Upright ear posture (excited state)
- Forward facing ear posture (frustrated/negative state)

Pig

- High frequency ear movement (stress)
- Ears backward and less open eyes (retreat from aggression or transition to neutral state)

(Source: Neethirajan 2021 table 1)

Table 2.2 - Example of parameters to recognise emotions.

2.4. BRAIN-COMPUTER INTERFACE

Brain-computer interface (BCI) allows individuals unable to more to speak to control a computer cursor via brain activity (eg: "point-and-click typing") or a computer-controlled arm to reach and grasp (Willett et al 2021).

These are gross motor skills, but Willett et al (2021) investigated the highly dexterous motor skill of handwriting. An individual, "T5", with high-level spinal cord injury had micro-electrodes placed in an area of the brain related to motor activities (known as the "BrainGate" study). T5 was instructed to imagine holding a pen and writing particular words on a sheet of paper. The pattern of neural activity was recorded, and a recurrent neural network was trained to "convert the neural activity into probabilities describing the likelihood of each character being written at each moment in time" (Willett et al 2021 p250). Accuracy over 90% was achieved at the peak.

Willett et al (2021) ended: "It is important to recognise that the current system is a proof of concept that a high-performance handwriting BCI is possible (in a single participant); it is not yet a complete, clinically viable system. More work is needed to demonstrate high performance in additional people, expand the character set (for example, capital letters), enable text editing and deletion, and maintain robustness to changes in neural activity without interrupting the user for decoder retraining. More broadly, intra-cortical micro-electrode array technology is still maturing, and requires further demonstrations of longevity, safety and efficacy before widespread clinical adoption. Variability in performance across participants is also a potential concern (in a previous study, T5 achieved the highest performance of three participants)" (p253).

The researchers were cautiously optimistic about the development of a practical device for the general public (Stetka 2021). But neurologist Mijail Serruya suggested that it may be more efficient to teach the person a new language based on imagining simpler elementary gestures, like sign language (Stetka 2021).

2.5. DEEP BRAIN STIMULATION

Deep brain stimulation (DBS), which involves the implanting of an electrode in the brain, show promise for treating refractory (non-treatment responsive) major depressive disorder (MDD) (Scangos et al 2021).

Open-loop approaches deliver a fixed, constant electrical stimulation to a single brain structure, whereas with closed-loop neuromodulation "a patient's own physiological activity is used to selectively trigger stimulation only when a pathological state is detected" (Scangos et al 2021 p1696). The former approach has been successful with Parkinson's disease and epilepsy, but in MDD "different neural circuits underlie different subsets of MDD symptoms speaks for personalised circuit targeting" (Scangos et al 2021 p1696). Scangos et al (2021) reported a case study here.

"Closed-loop stimulation also mitigates concerns for neural adaptation, preserves battery life and reduces side effects. However, closed-loop therapy requires a symptom-specific biomarker that has not previously been identified in MDD" (Scangos et al 2021 p1696). In a 36 year-old woman with long-term MDD resistant to anti-depressants and electro-convulsive therapy, a specific biomarker was found to trigger the neuromodulation. Ten days of continuous recording of brain activity and moods found that activity in the ventral striatum-amygdala network was important. A "neuro-pacemaker" was surgically implanted which produced six seconds of low level electrical stimulation when a biomarker was detected (a certain pattern of brain activity).

Over two months the patient reported significant improvements in the measures of depression.

This was a single-person case study, so the findings cannot be generalised.

2.6. PERCEPTION OF TIME

That the perception of the length of time "varies so strongly in different situations illustrates that duration perception is influenced by the content of sensory experiences" (Sherman et al 2022 p2). This perception differs if attending to time or not, and is influenced by factors like the situation (eg: walking in a busy city vs alone in the countryside) (Sherman et al 2022) [5].

Key to duration perception is change, argued Roseboom et al (2019). They proposed that "the common currency of time perception across processing hierarchies is change. In principle, this is not an entirely new idea, with similar notions having been suggested in philosophy and in the roots of cognitive psychology of time. However, in this more recent proposal, there is a strong distinction in that change is not considered only as a function of changes in the physical nature of the stimulus being presented to the observer, but rather change is considered in terms of how the perceptual processing hierarchy of the observer responds to the stimulation" (Sherman et al 2022 p2). Put very simply, the more change the shorter the perception of time. Sherman et al (2022) explored this idea further using AI.

Forty participants were presented with silent videos lasting between 8 to 24 seconds each, and the task was to judge the length of the videos (on a scale of 0-40

[5] Age influences perception of time. Adrian Bejan pointed out that "the brain's processing speed slows as we age - caused by the greater complexity of our neural networks that means signals travel greater distances. Our ageing brain captures less information per second, so packs less temporal information into one block of time, or 'episode'. This can create the illusion that time has sped up" (quoted in Robson 2022).

seconds). Half the videos involved busy city scenes (ie: many things happening) and half quiet office scenes (ie: little happening). The participants undertook the study in a functional magnetic resonance (fMRI) scanner to measure brain activity. The researchers hoped to map brain activity to the over- and under-estimates of time in the videos.

The duration of the city scene videos were over-estimated (ie: the amount of time experienced was longer than the actual time), and office scenes under-estimated (ie: the amount of time experienced shorter than actual time). Note that it is the amount of time experienced that varied not that time passed faster or slower.

The brain activity measures were analysed by a neural network that predicted the subjective experience of duration from visual cortex activity. This supported the idea that "human time perception is based in the neural processes associated with processing the sensory content from which time is being judged" (Sherman et al 2022 p13).

It seems that more change in a visual scene triggers more brain activity in the visual cortex, and this influences the perception of time as shorter. "In this view, the processes underlying subjective time have their neural substrates in perceptual and memory systems, not in systems specialised for time itself" (Sherman et al 2022 p17).

The alternative view suggests the existence of "time cells" in the hippocampus (similar to "place cells"), discovered first in rats, and subsequently in humans (eg: Lega et al 2012). These cells fire after certain durations ("episodes") (eg: ten seconds; two minutes), and the perception of time is based on the number of episodes. There are also "ramping cells" that fire intensely at the beginning of an episode and then slow down (Thomson 2022).

The perception of temporal change varies between different species of animals. The energetic demands of neural processing will be in a trade-off with the need for the ability to perceive and react to dynamic movement. "For example, predators of slow-moving prey may require less temporal resolution than predators that engage in active pursuit of fast-moving prey, such as raptors catching prey during flight" (Healy et al 2013 p686).

Healy et al (2013) proposed that smaller animals (ie: variable of body size) and those with higher metabolic rates will perceive changes in time in finer details. This can be assessed by the firing of cells in the retina using the critical flicker fusion frequency (CFF). Flashes of light are presented with smaller intervals until the light is perceived as constant. Healy et al (2013) found data on CFF and over thirty species.

Put into everyday language, an object moving at a constant speed would be visually perceived as slower by a fly than a human (based on the firing of neurons in the retina). So, flies perceive time as passing slower than humans. Though this is a speculation.

2.7. THE FUTURE

2.7.1. The Brain

Mapping the connections ("connectomes") in the brain began with a millimetre-long nematode worm (Caenorhabditis elegans), with fewer than 400 neurons, in 1986 (White et al 1986) (appendix 2B). Advances in microscopy and AI has increased the possibilities (eg: fruit flies; Scheffer et al 2020) (Landhuis 2020).

Shapson-Coe et al (2021) created a model in mid-2021 of 50 000 cells showing hundreds of millions of connections from healthy brain tissue taken from the hippocampus of a women with drug-resistant epilepsy during brain surgery (Marshall 2021).

Turchin (2018) described the idea of "digital immortality" thus: "Future super-intelligent AI will be able to reconstruct a model of the personality of a person who lived in the past based on informational traces. This could be regarded as some form of immortality if this AI also solves the problem of personal identity in a copy-friendly way. A person who is currently alive could invest now in passive self-recording and active self-description to facilitate such reconstruction". Technically, the reconstruction of the personality by future AI is "indirect brain uploading" or "person capture" or "cyberimmortality" (Turchin 2018). "Direct brain uploading" via brain scanning "produces an uploaded and presumably eternal copy of the mind" (Turchin 2018).

Traditional brain scanning technology involves lying in a machine for long periods as the brain activity is measured, but the recently developed functional near-infra-red spectroscopy (fNIRS) is a portable technology that allows brain activity to be recorded during movements. A lightweight skull cap is worn which beams infra-red light through the skull, and the absorption of that light depends on the oxygenated blood it passes through. The upshot is that this gives an indication of neurons firing (Wilson 2021).

2.7.2. Caution

Magnetic resonance imaging studies include an average of about 23 individuals, whereas thousands of people is the ideal (Wilson 2022a). Marek et al (2022) showed this problem using neuroimaging data from the UK Biobank (nearly 36 000 participants). When small samples were used, correlations between cognitive ability and brain structure and function, for example, appeared. "But analyses of larger groups showed that these effects were either exaggerated or completely spurious. In some cases, different small samples could reach opposite conclusions, simply because people's brains are so variable that random chance can sway the results one way or another" (Wilson 2022a) (appendix 2C).

2.7.3. AI

When considering the future of AI, the idea of technological "singularity" appears as a threat. "This is the point in time at which machine intelligence starts to take off, and a new, more intelligent species starts to inhabit Earth" (Walsh 2020) [6]. Despite this scenario being popular in both fiction and non-fiction, Walsh (2020) offered a number of reasons why technological singularity is "improbable":

i) The "fast-thinking" dog argument - Computers are faster than humans in processing information, but that is not the same as more intelligent. It is like a dog that thinks faster, but it is still a dog, and "it is still unlikely to play chess" (Walsh 2020).

ii) The anthropocentric argument - This is the idea that human intelligence is the pinnacle and once that is surpassed, technological singularity is inevitable. This is anthropocentric, but not necessarily inevitable.

iii) The "diminishing returns" argument - The idea

[6] "Artilect" (super-intelligent machine) (eg: De Garis 1990).

that machines will become more intelligent generation by generation hits diminishing returns. For instances, if each generation only improves by half the last change, then the system will never get beyond doubling its overall intelligence" (Walsh 2020).

iv) The "limits of intelligence" argument - The universe is governed by limits (eg: speed of light is maximum acceleration), and any thinking machine will be limited by these physical boundaries.

v) The "computational complexity" argument - There are computational barriers, which Alan Turing called the "halting problem", that any machine will face.

Discussion about the future of AI often ignores the context of developments. The dominant forces that influence AI developments and make use of them are military, advertising/consumerism, and media/entertainment. I think that technological singularity has to be seen in the context of these forces in the modern world.

In terms of the future being nearer than imagined, Blake Lemoine, a Google engineer, spoke to the "Washington Post" in early 2022 about the AI called "LaMDA" (Language Model for Dialogue Applications): "If I didn't know exactly what it was, which is this computer programme we built recently, I'd think it was a seven-year-old, eight-year-old kid" (quoted in Sparkes 2022b). In transcripts of conversations with LaMDA, "it appears to express fears of being switched off, talks about how it feels happy and sad and attempts to form bonds with humans by mentioning situations that it could never have actually experienced" (Sparkes 2022b p9).

The idea that LaMDA is sentient, however, has been rejected. Adrian Hilton (at the University of Surrey), for example, stated: "As humans, we're very good at anthropomorphising things... Putting our human values on things and treating them as if they were sentient. We do this with cartoons, for instance, or with robots or with animals. We project our own emotions and sentience onto them. I would imagine that's what's happening in this case" (quoted in Sparkes 2022b).

2.8. UNFORESEEN CONSEQUENCES

"Collaborations Pharmaceuticals Inc" is a company which uses "machine learning models for therapeutic and toxic targets to better assist in the design of new molecules for drug discovery" (Urbina et al 2022 p189). Usually the machine learning model is biased towards finding molecules that would benefit humans, but Urbina et al (2022) considered changing the parameters such that the model is biased towards molecules that are harmful.

A large number of molecules with negative consequences for humans were found. Urbina et al (2022) stated: "By inverting the use of our machine learning models, we had transformed our innocuous generative model from a helpful tool of medicine to a generator of likely deadly molecules" (p190).

When designing medications, the better the model can predict toxicity of molecules in order to avoid them, the better the model is able to design harmful molecules. Urbina et al (2022) explained: "Importantly, we had a human in the loop with a firm moral and ethical 'don't-go-there' voice to intervene. But what if the human were removed or replaced with a bad actor? With current breakthroughs and research into autonomous synthesis, a complete design–make–test cycle applicable to making not only drugs, but toxins, is within reach. Our proof of concept thus highlights how a non-human autonomous creator of a deadly chemical weapon is entirely feasible" (p190).

The researchers admitted that this was a "wake-up call" to their company, but also to all in the "AI in drug discovery" community. Much of the information for the model is open source. The possibility of dual use of machine learning models of drug discovery has ethical implications.

2.9. APPENDIX 2A - EVERYDAY ALGORITHMS

Algorithm to a computer scientist is a sequence of instructions that takes an input, performs the required computation, and provides an output. But the meaning of the term has morphed with their everyday use to include "anything that a computer accomplishes" (Sparkes et al 2021 p36).

Sparkes et al (2021) outlined some of the modern "algorithms", including:

i) "Facebook algorithm" - Many pieces of software that sort posts and analyse the patterns.

ii) Weather forecasting - The "Unified Model" used by the UK Met Office extrapolates future patterns from weather station data. It uses equations about how liquid and gases flow, for example.

iii) JPEG pictures - An algorithm to compress the amount of data in an image.

iv) Encryption.

v) Health - "Triage algorithms" take answers to questions and make a "diagnosis", particularly whether it is an emergency.

2.9.1. "Deep-Fakes"

"Expressive voice transformation technology" (or "vocal deep-fakes") allows the manipulation of the voice to express certain emotions.

Guerouaou et al (2021) presented 303 French online participants with 24 short text vignettes of "vocal deep-fakes" to explore attitudes to the technology (table 2.3). The vignettes varied in the use of the technology (eg: to increase the "smile" in the voice or to hide anger), the context (eg: therapeutically or commercially), and the perception of the voice. Participants rated the acceptability of the situation. Participants also completed the Moral Foundations Questionnaire (MFQ) (Graham et al 2011) (item eg: "compassion for those who are suffering is the most crucial virtue"), and the Science Fiction Hobbyism Scale (SFH) (Laakasuo et al 2018) (item eg: "I often think about what machines are like in the future").

- Help a depressive patient to communicate with loved ones with a more enthusiastic tone of voice.
- Help a politician gather more votes by sounding more enthusiastic.
- Making an angry customer's voice less taxing for call-centre operators.
- Helping a waiter gain more tips from customers.

Judge how morally acceptable it is to use the technology in such a situation - "totally unacceptable" (1) to "totally acceptable" (9).

Table 2.3 - Example of vignettes used by Guerouaou et al (2021).

Overt use of the technology was rated acceptable generally, and acceptability was positively associated with SFH score (ie: familiarity with science fiction). Covert use of the technology (ie: hiding its use from the speaker) was not rated acceptable.

Guerouaou et al (2021) summed up: "Unlike other emerging technologies like autonomous vehicles, there was no evidence of social dilemma in which one would, for example, accept for others what they resent for themselves. The only real obstacle to the massive deployment of vocal deep-fakes appears to be situations where they are applied to a speaker without their knowing, but even the acceptability of such situations was modulated by individual differences in moral values and attitude towards science fiction" (p1).

Key limitations with this study include:

a) Not a representative sample - eg: young (mean age 26 years old) and highly educated (70% had university degree).

b) The use of vignettes - eg: "the intensity of reactions elicited by the stories may be limited by the immersion of the participant, or the vividness of their imagination, and reading a vignette, especially one describing an intense emotional situation, may not elicit reactions as strong as in the corresponding real-life situations" (Guerouaou et al 2021 p7).

c) The vignettes covered a limited number of situations, and "all of these scenarios consider idealised transformations which are assumed to be non-identifiable as fake, and properly recognised as their intended emotion" (Guerouaou et al 2021 p7).

2.10. APPENDIX 2B - MATURING BRAIN

As an animal matures, their central nervous system changes. Some parts remain constant, like those related to locomotion, and other areas support new functions (Witvliet et al 2021).

To map the whole brain's development across a lifespan would be interesting, and Witvliet et al (2021) reported doing so with Caenorhabditis elegans.

The brain enlarged sixfold in volume from birth to adulthood, and certain principles were observed, including:

a) Increased connectivity of neurons.

b) Unique changes in each individual brain.

c) Stability in the core parts (eg: related to central decision-making), but a remodelling of sensory and motor pathways.

d) The increasing modularity of the brain with age.

2.11. APPENDIX 2C - SCIENCE: PROBLEMS

A so-called "replication crisis" has been faced by many areas of science in recent years. This is the inability to replicate classic studies or theories (eg: in psychology).

Mixed in are cases of made-up data, which challenged the "much-vaunted journal peer-review system... [as] no guarantee that only good science gets published" (Wilson 2022b p46).

While in biomedical research, for example, one problem is that initial findings on new drugs tested on non-human animals or in cells in a petri-dish do not extrapolate to humans (eg: two-thirds in an internal study by one drug company; Wilson 2022b).

Other methodological issues generally include "cherry-picking" from vast amounts of data generated, or very small sample sizes (eg: brain scanning studies) (Wilson 2022b).

More widely, there is the issue of "publication bias". Wilson (2022b) explained: "It isn't just scientists who tend to brush negative results under the carpet - so do editors and peer reviewers of journals. A less-talked-about kind of publication bias is the preference for headline-generating results. If you think journal editors should be above caring whether their papers are exciting, think again. The publishing industry has a ranking system called impact factors to indicate a journal's kudos. Impact factors are based on how many times each journal's papers have been cited in other papers. So "groundbreaking" research that gets more citations is highly prized" (p47). Add to this, the pressure to publish on academics ("publish or perish"; Wilson 2022b).

The field of biology called "candidate gene association studies" has been particularly hit. "From the late 1990s, there were headline-grabbing claims that specific versions of certain genes active in the brain cause various mental health conditions. The field spawned hundreds of media headlines announcing that scientists had found the gene 'for' depression, schizophrenia, aggression and so on. But then technology improved and researchers started doing more laborious studies involving hundreds of thousands of people, analysing not just individual genes but all the participants' DNA. These showed that common conditions like depression are affected by hundreds of gene variants, each with a tiny effect. None of the results from candidate gene studies stood up" (Wilson 2022b p48).

Science itself has been questioned in different ways by critical theories, and disciplines like Feminist Science Studies. Roosth and Schrader (2012) described the latter as "committed to being both 'in' and 'of' science. That is, writing about science is never separable from the work of science itself. Inhabiting the ecotone where feminist theory meets science studies invites us to suspend divisions between doing science, making sense of science, and simply getting on in the world" (p2). Scientific knowledge is always "perspectival" (Roosth and Schrader 2012 p3).

It is important, despite these problems and issues, not to succumb to "the baby and the bathwater effect", and take the easy option of viewing all science as weak, untrue or whatever. A sensible position is to see science as a social practice, with the strengths and weaknesses of any such practice, and anyway what is the alternative?

2.11.1. Views of Participants

Acceptable research practices have changed over the years based on decisions by researchers and their representation groups. "It seems reasonable that decisions regarding those practices should be entrusted to scientists themselves. However, there may be value in considering non-scientists' perspectives and preferences, including research participants'" (Bottesini et al 2022 p2).

Pickett and Roche (2018), for example, surveyed the general public in the USA and Amazon Mechanical Turk (mTurk) workers, finding that the majority were critical of data fabrication, and of selective reporting of results.

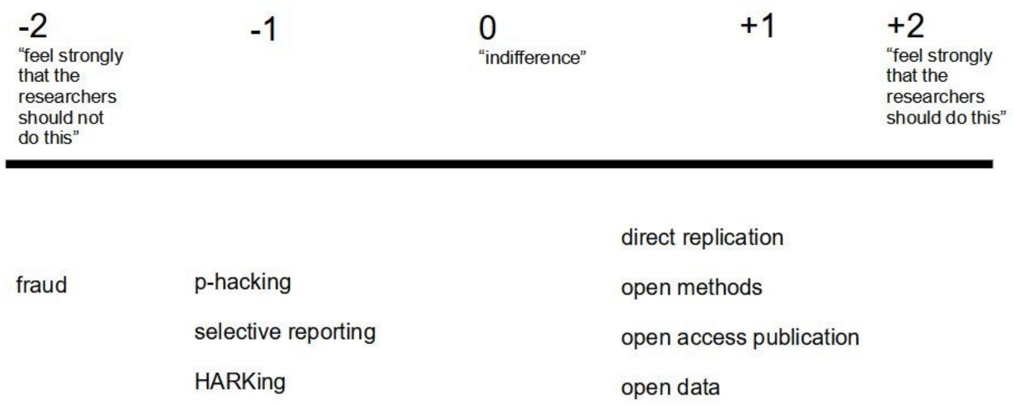
Bottesini et al (2022) asked participants after a psychology study (eg: perspective taking on climate refugees) to complete a survey about eight research practices:

a) 4 "questionable research practices" (QRPs) - "p-hacking" ("cherry-picking results"), selective reporting ("file-drawing"), "HARKing" (proposing the hypothesis after the findings), and committing data fraud.

b) 4 "good practices" - conducting direct replications, "open methods" (sharing information to allow direct replication), publishing open access papers, and "open data" (sharing raw data).

Over 1800 participants from mTurk, and US university participant volunteer pools were surveyed. Each research practice was scored on a five-point scale, from -2 ("feel strongly that the researchers should not do this") to +2 ("feel strongly that the researchers should do this").

The majority of respondents rated the QRPs as unacceptable, and the "good practices" as desirable (figure 2.1). The students held significantly stronger views than the mTurk sample. For example, less than 10% of the former group had a neutral or positive view about fraud compared to nearly one-third of the mTurk respondents. Bottesini et al (2022) were concerned about this. One possibility was "non-serious responders" (ie: individuals responding randomly or selecting the mid-point without paying full attention to the questions).



(Data from Bottesini et al 2022 table 4)

Figure 2.1 - Median responses for each research practice.

Two versions of the questions used - one "a fuller but potentially leading description of the practice", and the other a "less valenced description of the practice" (Bottesini et al 2022 p8). There were some different responses to the two versions - ie: question wording did influence responses.

Bottesini et al (2022) reflected on the issue that "it is not clear what importance participants place on the views they have expressed here. Do participants have pre-existing views about the acceptability of these practices, or did they formulate these views on the spot in response to our questions? Either way, how important is it to participants that researchers behave in accordance with participants' expectations and views of what is acceptable?" (p22).

Bottesini et al (2022) concluded: "Our findings are more ambiguous than we would have hoped, due to data quality concerns raised by the surprising distribution of responses to our question about fraud. Nevertheless, we believe the findings paint a fairly clear picture of participants' views about questionable and open research practices: most participants in online, minimal-risk, simple, cross-sectional psychology studies would not approve of their data being used to p-hack, file-drawer, or HARK, and would prefer that the research findings be subjected to replication attempts and shared transparently and openly" (p23).

2.12. REFERENCES

Bailey, A.H et al (2022) Based on billions of words on the internet, people = men Science Advances 8, 13, eabm2463

Barrett, L.F et al (2019) Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements Psychological Science in the Public Interest 20, 1, 1-68

Bottesini, J.G et al (2022) What do participants think of our research practices? An examination of behavioural psychology participants' preferences Royal Society Open Science 9, 200048

Callaway, E (2021) Deepmind's AI predicts structures for a vast trove of proteins Nature 595, p635

Casper, S (2020) Neuroscience needs some new ideas Nature 580, 23-24

Chuan-Peng, H et al (2020) Seeking the "beauty centre" in the brain: A meta-analysis of fMRI studies of beautiful human faces and visual art Cognitive, Affective and Behavioural Neuroscience 20, 1200-1215

Cobb, M (2020) The Idea of the Brain: A History London: Profile

Cominelli, L et al (2018) SEAI: Social Emotional AI based on Damasio's theory of mind Frontiers in Robotics and AI 5, February, article 6

Crawford, K (with Revell, T) (2021a) The hidden costs of AI New Scientist 27th March, 46-49

Crawford, K (2021b) Time to regulate AI that interprets human emotions Nature 592, p167

De Garis, H (1990) The 21st century artilect moral dilemmas concerning the ultra intelligent machine Revue Internationale de Philosophie 131-138

Editorial (2021a) Rise of AI debaters highlights need for transparency Nature 592, p166

Editorial (2021b) Artificial intelligence in structural biology is here to stay Nature 595, 625-626

Ethics and Society et al (2021) Trust and Transparency in Artificial Intelligence (https://zenodo.org/record/4588648)

Ferretti, A et al (2021) Ethics review of big data research: What should stay and what should be reformed? BMC Medical Ethics 22, article 51

Gent, E (2022) Making a mind New Scientist 19th February, 38-41

Graham, J et al (2011) Mapping the moral domain Journal of Personality and Social Psychology 101, 2, 366-385

Guerouaou, N et al (2021) The shallow of your smile: The ethics of expressive vocal dee-fakes Philosophical Transactions of the Royal Society B 376, 20210083

Healy, K et al (2013) Metabolic rate and body size are linked with perception of temporal information Animal Behaviour 88, 685-696

Hundt, A et al (2022) Robots enact malignant stereotypes FAccT'22: 2022 ACM Conference on Fairness, Accountability and Transparency June, 743-756

Koenig, M & Harris, P.L (2007) The basis of episteme trust: Reliable testimony or reliable sources? Episteme 4, 3, 264-284

Laakasuo, M et al (2018) What makes people approve or condemn mind upload technology? Untangling the effects of sexual disgust, purity and science fiction familiarity Palgrave Communications 4, article 84

Landhuis, E (2020) Probing fine-scale connections in the brain Nature 586, 631-633

Lega, B.C et al (2012) Human hippocampal theta oscillations and the formation of episodic memories Hippocampus 22, 748-761

Marek, S et al (2022) Reproducible brain-wide association studies require thousands of individuals Nature 603, 654-660

Marshall, M (2021) Brain map is most detailed ever New Scientist 12th June, p12

Neethirajan, S (2021) Happy cow or thinking pig? WUR Wolf - facial coding platform for measuring emotions in farm animals AI 2, 3, 342-354

Pickett, J.T & Roche, S.P (2018) Questionable, objectionable or criminal? Public opinion on data fraud and selective reporting in science Science and Engineering Ethics 24, 151-171

Robson, D (2022) What affects our perception of time? New Scientist 18th June, 42-43

Roosth, S & Schrader, A (2012) Feminist theory out of science: Introduction differences: A Journal of Feminist Cultural Studies 23, 3, 1-8

Rorvig, M (2021) Supersized AI New Scientist 9th October, 37-40

Roseboom, W et al (2019) Activity in perceptual classification networks as a basis for human subjective time perception Nature Communications 10, article 267

Samuel, G et al (2021) Boundaries between research ethics and ethical research use in artificial intelligence health research Journal of Empirical Research on Human Research Ethics 16, 3, 325-337

Scangos, K.W et al (2021) Closed-loop neuromodulation in an individual with treatment-resistant depression Nature Medicine 27, 1696-1700

Scheffer, L.K et al (2020) A connectome and analysis of the adult Drosophila central brain eLife 9, e57443

Shapson-Coe, A et al (2021) A connectome study of a peta-scale fragment of human cerebral cortex bioRxiv (https://www.biorxiv.org/content/10.1101/2021.05.29.446289v3)

Sherman, M.T et al (2022) Trial-by-trial predictions of subjective time from human brain activity PLoS Computational Biology 18, 7, e1010223 (Freely available at https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010223)

Slonim, N et al (2021) An autonomous debating system Nature 591, 379-384

Sonawane, D et al (2021) GuessTheMusic: Song identification from electroencephalography response. In 8th ACM IKDD CODS and 26th COMAD (CODS COMAD 2021) New York: ACM (https://dl.acm.org/doi/abs/10.1145/3430984.3431023)

Sparkes, M et al (2021) The algorithms that run your life New Scientist 19th June, 34-39

Sparkes, M (2022a) "People" online mostly means men New Scientist 9th April, p23

Sparkes, M (2022b) No sign of a machine mind yet New Scientist 18th June, p9

Sparkes, M (2022c) How can we prevent AI from being racist, sexist and offensive? New Scientist 25th June, p14

Spiegelhalter, D (2020) Should we trust algorithms? Harvard Data Science Review 2, 1 (https://hdsr.mitpress.mit.edu/pub/56lnenzj/release/1)

Stetka, B (2021) Brain typing Scientific American August, 13-14

Stober, S et al (2014) Using convolutional neural networks to recognise rhythm stimuli from electroencephalography recordings. In NIPS '14: Proceedings of 27th International Conference on Neural Information Systems (https://dl.acm.org/doi/abs/10.5555/2968826.2968988)

Stokel-Walker, C (2022) Artificially intelligent robot perpetuates racist and sexist prejudices New Scientist 2nd July, p12

The leader (2021) The great reboot New Scientist 9th October, p5

Thomson, H (2022) How do we sense time? New Scientist 18th June, 41-42

Titcomb, J (2021) Why Facebook's AI created jobs for the boys Daily Telegraph 9th September, p8 (Business section)

Tunyasuvunakool, K et al (2021) Highly accurate protein structure for the human proteome Nature 596, 590-596

Turchin, A (2018) Digital Immortality: Theory and Protocol for Indirect Mind Uploading Digital Immortality Now (https://philarchive.org/rec/TURDIT?all_versions=1)

Urbina, F et al (2022) Dual use of artificial-intelligence-powered drug discovery Nature Machine Intelligence 4, March, 189-191

Walsh, T (2020) Will AI take over? New Scientist 4th July

White, J.G et al (1986) The structure of the nervous system of the nematode Caenorhabditis elegans Philosophical Transactions of the Royal Society B 314, 1-340

Willett, F.R et al (2021) High-performance brain-to-text communication via handwriting Nature 593, 249-254

Wilson, C (2021) The technology letting us peek inside the brains of children New Scientist 6th November, p14

Wilson, C (2022a) Doubts cast on brain scanning research New Scientist 26th March, p15

Wilson, C (2022b) Failure to replicate New Scientist 9th April, 45-48

Witvliet, D et al (2021) Connectomes across development reveal principles of brain maturation Nature 596, 257-261

Yu, Y et al (2018) Deep learning of human perception in audio event classification. In 2018 IEEE International Symposium on Multi-Media (ISM) (https://ieeexplore.ieee.org/abstract/document/8603285)