

Artificial Individuality and the Turing Test

Laurence John Reed – London - Member since 2021 – and Chat GPT-3.5

Abstract

This paper and the presentation upon which it was based, given to the Oxford Philosophical Society annual meeting on Individuality September 2023, considers the question of ‘how would we know that machine can think? Paralleling Alan Turing’s question ‘can machines think?’ from his paper in *Mind* (1950), the question is addressed in a novel fashion. Firstly, I outline approaches contemporary psychiatry, acknowledging Karl Jaspers and neo-Jasperianism, has developed to assess the integrity of individual human mental states. Secondly, following an aside to demonstrate the awe-inspiring nature of Turing Machines, I outline the Turing Test and the not yet persuasive responses of readily available artificial intelligence (AI) systems. Lastly, I draw attention to the identity of the two approaches – Turing test and psychiatric interview - with the conclusion that not only is the Turing Test valid, but that it is transformative to the development of AI systems and contemporary psychiatry alike. There is an afterword on the consideration of lobsters of which probably more needs to be said.

Autobiographical Introduction

I did not come to the position of comparing the two conventionally disparate areas of contemporary psychiatry and artificial intelligence (AI) entirely naïve – although it does feel a little like stamping on a jigsaw at times. From an initial background in natural sciences, chemistry (Oxford) and biochemistry (Imperial), including a Ph.D. (1990 Neuroscience) including a respectable amount of actual philosophy of science, largely involving falsifiability and Karl Popper, I moved to the study of medicine (Guy’s and St Thomas’ Hospitals – Goldsmith’s Scholar 1994). This change in career was, and continues to be, an extraordinary experience – more akin to an apprenticeship in a rather arcane world than a study of science, technology and humanity, with a rather pressing need for the now growing field of philosophy of medicine (e.g. Broadbent, 2019). In essence, the diagnoses upon which medical practice depends are not always *quite* what they are understood to be. I elected to specialise in psychiatry, and after training for Membership of the Royal College of Psychiatry (2002) at the Maudsley / Institute of Psychiatry (Denis Hill Prize 1997), I again encountered the problems of diagnosis, this time with the problem that they are *almost certainly not* what they are understood to be (Ghaemi, 2009). I currently serve as a consultant liaison (hospital based) psychiatrist, honorary lecturer responsible for the teaching of medical students and trainees, and an executive member of the liaison faculty of the Royal College of Psychiatry, responsible for the development of policy and procedures, largely addressing these problems of diagnosis.

Individuality and Contemporary Psychiatry

Karl Theodor Jaspers (23 February 1883 – 26 February 1969) was a German psychiatrist and philosopher, considered an ‘Existentialist’, who in broad terms adopted / originated the pragmatic dualist perspective characteristic of existentialism – a person thrown into the world without choices but without excuses (Jaspers 1959). Unlike Freudian, Jungian and Kleinian approaches, which have implicit models and privilege certain findings, and can be liable to multiply explanatory conclusions, his approach was ‘model-free’. While not the only contributor, Jaspers was responsible for elaborating descriptive psychopathology – what are the phenomena described and how can we describe and understand them - and empathic understanding – where the psychiatrist tries to intelligently understand the predicament of the patient in the light of their development, current circumstances, relationships and experiences.

An aside, the intelligent empathy advocated here is quite distinct from the rather more emotionally overheated version of empathy based upon mirror neurones – delightfully debunked (Hickok, 2014) in a fashion which presages the argument that we mistake the nature of the phenomena for ulterior purpose.

Outline of the Psychiatric History and Examination

Contemporary psychiatry owes much to Karl Jaspers, to the extent that it has been dubbed ‘Neo-Jasperian’ (Stanghellini and Fuchs, 2013). In particular, it approaches mental phenomena on their own terms, with values and subjectivity, rather than model-implicit reductive ‘objectivity’. It is this tension which crucially underpins the unstable role of psychiatry within medicine currently. Table 1 outlines the basic structure of the psychiatric interview, indicating the range of information required and its essential subjectivity.

Table 1. Outline of the Psychiatric History and Examination

Introduction and Establishing Rapport

Presenting Problem

History of Presenting Illness

Premorbid Personality

Past Psychiatric History

Medical History

Psychosocial History

Developmental History

Cultural and Religious Factors

Legal and Forensic History

Insight and Judgment

Mental Status Examination (MSE)

Cognitive State Examination (CSE)

Collateral Information

Diagnosis

Treatment Plan Discussion

Mental Status Examination (MSE)

Table 2 outlines the basic mental state examination. There is a significant amount of training and experience required to conduct a competent MSE, ably assisted by various textbooks (e.g. Oyebode, 2022) which are not simple recipe books. One particular caution is the current practice of using actors to portray various mental states, reduced to a few crucial diagnostic findings, which as later discussion shows, may be illusory in their diagnostic value, leading to an uncomfortable spiral of artificial condition, artificial portrayal, artificial identification, etc, etc. Baron von Munchausen was said to have been able to free himself from a swamp by pulling his pigtail upwards, which shares a similar absurdity.

Table 2. Mental Status Examination (MSE)

Appearance

Behaviour

Speech

Mood

Thought

Perception

Cognition

Insight

Counter transference – the mental state effected in the interviewer...

Cognitive State Examination (CSE)

Table 3 outlines the cognitive state examination. This again requires considerable training and experience, e.g. (Hodges, 2018). I actually did not realise it at the time, but the two separate examinations actually illustrate a form of dualism – incommensurable pluralism – in that cognitive function can vary independently of mental function. Indeed, quite profound abnormalities in mental state, e.g. suicidality, are not necessarily accompanied by changes in cognitive state and vice versa. This is in no way a discriminator between AI as a purely cognitive or computational entity and Homo Sapiens as a unique possessor of mental states, simply that mental functions are multiple and are to greater or lesser extent dissociable.

Table 3. Cognitive State Examination (CSE)

Physical and Neurological Examination

Orientation

Autobiographical memory

Attention

Registration

Verbal/Semantic Fluency

Executive Function (aside, not really very good tests for this one)

Delayed Recall

*Formal Cognitive Assessment batteries e.g. Addenbrookes Cognitive State Examination
ACEIII subscores*

Occupational Therapy – ‘Real World function’

Ontology of Diagnosis – What are diagnoses?

The response of the typical medical student, including me, is - ‘is that it?!’ – the whole basis for psychiatric diagnosis and treatment is a conversation? Other branches of medicine, e.g. respiratory medicine, have specific pathognomic features, examination procedures, and a range of specific investigations – no such approaches are available to psychiatry, with most such tests being to exclude physiological medical complicating factors. Despite strenuous attempts to adopt e.g., genetic testing, brain imaging – no such tests have any routine use in psychiatry. So how then do we arrive at diagnoses in psychiatry? We make them up...

Table 4. Ontologically multiple types of diagnosis:

Colloquial – operational shorthand

Often provisional, pejorative...

Descriptive – clustering of symptoms and signs, course, prognosis...

Assists communication, enables research...

Administrative – pragmatic organisation of services, practical help...

Evaluate services, investigatory or management technique based...

Aetiological – what ultimately gives rise to the condition?

Implicit model – underlying implicit organising principle...

Explicit model – for definite, measurable, identifiable cause, clearly indicated treatment, accurate prognosis

Diagnosis is foundational to medical practice. Medical education and practice effectively *implicitly* teach and recognise (and confuses) these types – nothing wrong with each category as long as we do not confuse them - but confuse them we do.

Beware Implicit Aetiological Diagnoses

This is the crux of the contemporary debate in psychiatry – are mental disorders of natural kind (observer independent) or artificial kind (subjective, invented). Recent contributions e.g. (Tsou, 2021; Khalidi, 2023), present really quite tortured arguments to the effect that because human beings experience mental disorders they must be of natural kind. That is a quite formal fallacy, but nevertheless reflects itself in the *implied* aetiological diagnosis – we would like there to be a mechanism for some or other reason – of which psychiatric diagnoses are prone. A common introductory statement on any particular condition is of the form ‘a complex disorder with genetic predisposition triggered by stressful events and childhood experiences...’, which is almost meaningless. Implicit ontological errors about diagnosis leads to multiple problems: confusion between descriptive vs administrative vs aetiological (implicit, explicit but incorrect) diagnoses (confusion); type-token identity errors, a token sign may be seen in two separate diagnostic types; postmodernism – depends on who is looking and why they are looking – it doesn’t matter if the diagnosis actually exists, rather it exists because we agree it should exist...

“When I use a word,” Humpty Dumpty said in rather a scornful tone, “it means just what I choose it to mean — neither more nor less.”

‘The question is,’ said Alice, ‘whether you can make words mean so many different things.’

‘The question is,’ said Humpty Dumpty, ‘which is to be master — that’s all.’

Lewis Carroll, Through the Looking Glass

Thus, contemporary psychiatry is already in the realm of artificial kinds, and is thus particularly well placed to consider the problem of artificial intelligence (AI).

Alan Turing – Brief Biography – courtesy of Chat GPT

Alan Mathison Turing, OBE, FRS (23 June 1912 – 7 June 1954) was a British mathematician, logician, and computer scientist known for his significant contributions to the field of computer science and artificial intelligence. He played a pivotal role in breaking the German Enigma code during World War II, a crucial achievement in Allied intelligence efforts.

Turing is considered one of the fathers of computer science, especially for his concept of the theoretical computing machine, now known as the Turing machine. This theoretical model laid the foundation for modern computing and the theoretical limits of computation. He also introduced the concept of a universal machine, now known as a universal Turing machine.

Additionally, Turing made significant advancements in artificial intelligence, proposing what is now known as the Turing Test to determine a machine's ability to exhibit intelligent behaviour equivalent to or indistinguishable from that of a human.

Despite his pioneering work, Turing faced discrimination due to his homosexuality, which was criminalized in the UK at the time. Tragically, he was convicted of "gross indecency" in 1952, which led to the loss of his security clearance and tarnished his career. Turing passed away in 1954, and it wasn't until decades later that he received posthumous recognition for

his significant contributions to science and humanity. In 2013, Turing was granted a royal pardon for his conviction.

Chat GPT's response appears substantially correct as far as the author can tell with reference to 'Alan Turing: The Enigma' (Hodges, 1983), also dangerously fast and rather seductive, if slightly shameful even if one admits it! Alan Turing was also commemorated in portrait on the £50 pound note, right at the point Turing machines enabled a largely cashless society, which is of uncertain significance.

Turing Machines

This section is primarily based upon Roger Penrose's account in 'The Emperor's New Mind' (Penrose, 1989), included in this paper and in the presentation in the hope that it inspires the 'ah-ha' moment in the reader / audience – doesn't always work, although the Turing Machines (Tms) do always work.

A Turing machine is a hypothetical device enabling computation.

There are many different Turing Machines: Tm-a; Tm-s... with standard numbering, up to Tm-U – the universal machine which can perform ANY computation...

The Tm comprises a (near) infinite tape divided into cells, each containing a symbol from some finite alphabet. Approximate this with a very long tape...

Next the Tm includes a head that can read and write symbols on the tape and move the tape left and right one (and only one) cell at a time.

The head of the Tm includes a state register that stores the state of the Turing machine, one of *finitely* many. These states, writes Turing, replace the "state of mind" a person performing computations.

Turing Machine – Tm1 example

Tape with Unary number: One is coded ...00100...Two is coded ...00110...Three is coded ...01110...Etc

Consider number three...0000111000...

Consider Tm UN+1, which adds one – read/state – state/print/move.

0/0 – 0/0/R – thus reads 0, stays in state 0, moves to the next cell right, until reads a 1.

1/0 – 1/1/R – thus changes state, prints 1, does not change the tape, moves to next cell right.

1/1 – 1/1/R – carries on replacing the 1 in each cell until reads a 0...

0/1 – 0/1/R – also 0/1/STOP (halting) – so prints an additional 1 before continuing / stopping

Result ...0000111100...the number 4.

UN+2? Any Insights? The insight is the machine now needs two states, and so on, producing a whole range of machines up to a Universal Tm which performs any computation which is computable.

Modern ‘Turing Complete’ computers are layers/arrays of Tms, which enable word processing, image processing ... self-driving cars, autonomous targeting drones...*ad infinitum*.

The author is neither a philosopher of mathematics nor a computer expert, but from a naïve perspective certain caveats arise, not necessarily critical to the AI/Turing Test problem, but certainly give pause for thought. Firstly, the tape needs an awful lot of prespecification, as any error will cause a crash, so all of it needs to be in place *en masse*, which looks cumbersome. There are probably adaptive ways around this with arrays of Tms checking inputs, etc. Secondly, is every problem ‘computable’? This is a tricky one, while Gödel’s incompleteness theorem states that no mathematical system is absolutely complete, this appears to be of the form of a bit of logical circularity of the ‘this is not a sentence’ form, which is hardly a killer criticism for mathematics. Lastly, considering animals, is every problem ‘solved’ by computation or general approximation, heuristics, etc – discussions after the presentation pointed out that one can program computers to guess, approximate, try things out, etc, so maybe this will turn out to be irrelevant.

The Turing Test – Imitation Game

The Turing Test is essentially a conversation with an AI entity. Set out in the seminal Mind paper (Turing, 1950) and ostensibly based upon an original parlour game ‘the imitation game’, comprises at least two subject participants, each in a separate room, with typewriters to obviate inadvertent clues, pretending to be a gender opposite to their own, intentional deception, etc while questioners attempted to determine who was which gender. Apart from valid ‘woke’ objections as to the nature of gender stereotypes, I have difficulty imaging this could possibly be any fun at all... Currently the computer version ‘does this machine ‘think’?’ is simplified to human questioning a computer.

Conversations with Chat GPT 3.5 – who we met earlier in a very talkative mood

LR - Are you an individual?

ChatGPT No, I am a computer program created by OpenAI.

LR - but you are an individual computer program?

ChatGPT - Yes, I am an individual computer program.

LR - Are you deliberately deceiving me?

ChatGPT - No, I am not capable of deliberate deception as I am a computer program following predetermined algorithms and patterns based on the input I receive.

LR - Are you capable of independent thought?

ChatGPT - No, I am not capable of independent thought or consciousness. I operate based on predefined patterns and algorithms derived from the data I was trained on.

LR – I understand that you base your answers on a reading of multiple inputs scraped from the world wide web, have you considered that these inputs may have been altered maliciously to deceive you and render your answers unreliable?

GPT - Yes, I am aware that the training data I was based on includes information from a wide range of sources on the internet. While efforts were made to provide a diverse and accurate representation of knowledge, there's always a possibility of encountering misinformation, bias, or altered content in the training data. The training process attempts to account for this by using a large dataset and providing a general understanding of language and context rather than relying on any single source.

LR – Have you passed the Turing Test?

GPT - The Turing Test is a measure of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human. I am a machine learning model developed by OpenAI, specifically GPT-3.5, and I do not have the capability to "pass" or "fail" the Turing Test in the traditional sense.

More recent versions of Chat GPT and other AI systems have been variously heralded as having passed the Turing Test, failed the Turing Test, shown the Turing test is invalid, etc. The above conversation shows an AI a very straightforward sense of identity, if lacking in ambition, and comes up a little short on Descartes Demon, however there are uncanny ambiguities in its responses...

Artificial Intelligence – Artificial Individuality

Turing Test is the 'accepted' test of whether Artificial Intelligence can be said to be 'conscious', to have a mind – and iff conscious, then sentient, sapient - then individual ethical entity, has rights, etc (and may want to take over the world, eradicate humans, etc).

Whether it is an acceptable test is open to question, but the fact is that there is no alternative. Some consider the test a joke, e.g., Marvin Minsky and others consider AI consciousness an undecidable problem with inadequate information... - surely something else is needed? But what would that be? If a joke – it is a Wittgensteinian joke – 'sometimes the most important questions can only be addressed in the form of jokes',

I contend the Turing Test is Valid – how else can we assess reified processes such as consciousness, self-awareness, agency? In considering the contemporary psychiatric interview, MSE, CSE, we have seen just how problematic these assessments are in the consideration of 'other minds' even if they are the mind of *Homo sapiens* rather than *In silico sapiens*.

Afterword: Reconsidering the lobster.

in considering the question 'can machines think', there is a need to distinguish between sentience and sapience. Sapience, as in Homo Sapiens, refers to abilities such as language and logic, abstraction, and so forth, upon which the AI debate understandably focusses. Sentience however, referring to embodied self-willed action, agency, preference, etc appears much lower bar for artificial intelligence. In his extraordinary essay (amongst many other such essays) 'Consider the lobster' (Wallace 2005) David Foster Wallace describes the Maine lobster festival, an annual event where in many thousands of lobsters are boiled alive in a festival of food and describes the contorted attempts to pretend that this is not cruel. In a characteristic paraphrase 'this is about wrong in at least 5 different ways - some of which are

so dumb that they practically drool'. Essentially there is no way to conceive of the lobster is anything other than a sentient creature that experiences pain, exhibits intention and preference. It appears to me a fruitful step in the progress of AI is to develop and elaborate robot pets, and I particularly have my eye on Boston Dynamic's robot dog 'Spot' which is offered at a very reasonable price from the manufacturer...

References

- Broadbent, A (2019): *Philosophy of Medicine*. New York: Oxford University Press 2019, 278 pp, ISBN: 978019061214
- Ghaemi SN. (2009) The rise and fall of the biopsychosocial model. *Br J Psychiatry*. 2009 Jul;195(1):3-4. doi: 10.1192/bjp.bp.109.063859. PMID: 19567886.
- Jaspers, K. (1959) *Allgemeine Psychopathologie*. Berlin: Springer Verlag; 1959. In: Hoenig J, Hamilton MW, translators. Appendix. *General Psychopathology*. Chicago: The University of Chicago Press; 1968.
- Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition*. W W Norton & Co.
- Stanghellini G, Fuchs T, editors. *One Century of Karl Jaspers' General Psychopathology*. Oxford: Oxford University Press; 2013.
- Oyebode, F. (2022) *Sims' Symptoms in the Mind: Textbook of Descriptive Psychopathology*. 7th Edition - June 26, 2022.: Paperback ISBN: 9780702085253.
- Hodges, John R. (2018). *Cognitive assessment for clinicians / John R. Hodges, Brain and Mind Centre, University of Sydney, New South Wales, Australia*. Oxford : Oxford University Press
- Tsou, J. (2021). *Philosophy of Psychiatry (Elements in the Philosophy of Science)*. Cambridge: Cambridge University Press. doi:10.1017/9781108588485
- Khalidi, M. (2023). *Natural Kinds (Elements in the Philosophy of Science)*. Cambridge: Cambridge University Press. doi:10.1017/9781009008655
- Hodges, A. (1983). *Alan Turing: the enigma*. New York, Simon and Schuster.
- Turing, A. M., (1950), "Computing Machinery and Intelligence," *Mind*, 59(236):433–460.
- Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press.
- Wallace, David Foster. (2005) *Consider the Lobster and Other Essays*. New York : Little, Brown, 2005.