STEPHEN WEBB

EXPLORING AN EXTENSION TO DICK'S "INTELLIGENCE PRINCIPLE"

1. *Introduction* 2. *Critique of the Intelligence* 3. *The spread of intelligence*
4. *Ethical implications* 5. *Conclusions*

ABSTRACT: *EXPLORING AN EXTENSION TO DICK'S "INTELLIGENCE PRINCIPLE"*
Revisiting Dick's "Intelligence Principle" in the light of recent technological developments, I present an argument for why one might consider a slight extension to the Principle and I explore some of the implications of the suggested modification.

1. *Introduction*

In his paper *Cultural evolution, the postbiological universe and SETI*, the NASA historian Steven J. Dick adumbrated a theoretical construct he called the Intelligence Principle[1]. The Intelligence Principle states that «the maintenance, improvement and perpetuation of knowledge and intelligence is the central driving force of cultural evolution, and that to the extent intelligence can be improved, it will be improved».

Dick argued that, at the level of knowledge, this principle can be seen in daily operation: at both the individual and societal levels a common (though seemingly not universal) disposition is to attempt to maximise knowledge in order to gain advantage in the world.

[1] S.J. Dick, *Cultural evolution, the postbiological universe and* SETI, in «Int. J. Astrobiology», 2, 1, 2003, pp. 65–74.

Better education, improved technology, and increased access to information is commonly (although again not universally) seen as being advantageous to both individual and society.

Dick, however, was more interested in applying his principle at the species level. He argued that, for the species *Homo sapiens sapiens*, intelligence – which he asserted is a function of brain size and structure – has permitted it to outcompete other life forms and to dominate the planet. Brain size and structure, however, has not changed in 100,000 years. Dick argued the Intelligence Principle implies that, given the opportunity to increase intelligence – and, with recent developments in generative computing technology, one can readily appreciate how AI presents us with this opportunity[2]; biotechnology and genetic engineering are two other relevant technologies – any society would do so, or fail to do so at its peril. In this paper I provide a brief critique of the Intelligence Principle. Taking a different philosophical standpoint, however, I arrive at a similar conclusion to Dick. I propose an extension to the Intelligence Principle, and briefly explore the impact of this extension to the Fermi Paradox.

2. *Critique of the Intelligence Principle*

One can criticise the underlying assumptions and implications of the Intelligence Principle, as applied to biological creatures, and in particular to modern *Homo sapiens sapiens*, on several grounds. First, it lacks a solid grounding in empirical evidence. As Lineweaver[3] points out, when we use the paleontological record to plot a graph of brain size versus time, and then claim the results show a trend towards increasing intelligence (brain size acting as a proxy for intelligence), a selection bias is at work: we are

[2] See, e.g., Mckinsey & Co, *What's the future of generative AI? An early view in 15 charts*, www.mckinsey.com/featured-insights/mckinsey-explainers/whats-the-future-of-generative-ai-an-early-view-in-15-charts, 2023.

[3] C.H. Lineweaver, *Paleontological tests: human-like intelligence is not a convergent feature of evolution*, in J. Seckbach, M. Walsh M (eds), *From Fossils to Astrobiology*, Springer, Berlin, 2008, pp. 355–368.

plotting a property that defines humans. Similar trends would be apparent if one plotted the evolution of any extreme feature characterising a species (neck length in giraffes, nose length in elephants); such trends say nothing about life in general.

Second, and related to the first point, the Principle could be seen to downplay the role of chance and contingency. If one looks at the evolutionary record, for example, it becomes clear that for many species (e.g. *Homo sapiens sapiens*) an increase in sophistication has proven to be beneficial; for other creatures (e.g. parasitic flukes) there has been a pay-off in becoming simpler. Species respond to the pressures acting on them at each moment in time; there appears to be little evidence in the biological record to support Sagan's[4] dictum that at the species level «it's better to be smart than stupid». Indeed, Chomsky[5] has argued that the actions of humans since 1953 demonstrate the need to treat the dictum with suspicion.

Third, there remains a fundamental ambiguity in defining the concept of 'intelligence'. Bräuer, Hanus, Pika, Gray, and Uomini[6], for example, argue that the term 'cognition' has often been used by applying an anthropocentric rather than biocentric viewpoint, and that there is not "one cognition".

Fourth, the Principle takes relatively little account of the critical role played by cultural transmission and imitation, while aspects of the Principle could be argued as adopting a reductionist view of the inherently complex, multifactorial phenomenon of cultural evolution, a phenomenon that is shaped by social dynamics, environmental influences, historical contingencies, and economic

[4] C. Sagan, *The abundance of life-bearing planets*, in «Bioastronomy News», 7, 4, 1995.

[5] N. Chomsky, *Prospects for Survival*, Invited talk, UMass Amherst. Available from: www.nepm.org/podcasts-projects/2017-04-24/noam-chomsky-weighs-our-prospects-for-survival, 2017.

[6] J. Bräuer, D. Hanus, S. Pika, R. Gray, N. Uomini, *Old and new approaches to animal cognition: there is not 'one cognition'*, in «J. Intell.», 8, 3, 28, 2020.

systems. Fifth, it could be seen to possess an apparent teleological aspect. Finally, a Principle such as this has the potential for the (perhaps inadvertent) promotion of elitism.

Dick himself was aware of these critiques, but in his paper he was concerned less with the complexities of present-day cultural evolution than with long-term evolution. He argued convincingly that, when one considers cosmic timescales, one should adopt a "Stapledonian" mindset. Olaf Stapledon was an English philosopher concerned with the long-term prospects of the human species, and therefore with questions such as: How might our species evolve? What technological futures are most likely? What might humanity be capable of achieving? The adoption of such a mindset led Dick to postulate that biological and cultural evolution would necessarily lead to a postbiological universe, one in which corporeal intelligence cedes to artificial intelligence. The above critique of the Intelligence Principle necessarily requires modification if one assumes the existence of a postbiological universe; however, since the culture, drives and behaviours of postbiological beings are unknown, and perhaps from our vantage point unknowable, except that perhaps, unlike biological evolution, artificial intelligence is likely to be more explicitly goal-oriented and driven by design, I conclude the analysis at this point.

My personal viewpoint is that the phenomenon of intelligence, at the human level, arose through a myriad of chance events. Teleology plays no role. No evolutionary drive lead to *Homo sapiens sapiens* sitting atop a ladder of increasing sapience, and there is no guarantee that, in the short term, our intelligence (however we choose to define it) will be improved. Indeed, to echo Chomsky, it seems entirely possible that at the species level a combination of high intelligence and opposable thumbs is likely to lead to disaster. Nevertheless, for reasons given below, I concur with Dick regarding the importance of maintaining and perpetuating intelligence in the cosmos. This is an element of the Intelligence

Principle – but I argue that key to this is the spreading of intelligence throughout the cosmos.

3. *The spread of intelligence*

Recent work[7] provides evidence that, at around the time of the formation of our solar system, both type Ia and core-collapse supernovae seeded heavy elements into the cloud from which the planets condensed. Some 4.5 billion years later, labyrinthine chemical and biochemical processes, and the press of evolution, brought into being stable combinations of those elements possessing the quality of self-awareness. To apply another dictum of Sagan[8]: «we are a way for the universe to know itself». I believe this to be one of the most profound statements with philosophical content made by a practising astronomer.

One need ascribe no teleological aspect to Sagan's statement. Although science has yet to reveal many details regarding, among other things, abiogenesis, the emergence of prokaryotic life, and the evolution of high intelligence, we have no reason to suppose that the development of these characteristics involves anything other than the natural outcome of physicochemical processes operating in an environment that has remained relatively stable over aeons. Nevertheless, there is something wonderful – in the original sense of the word – in the notion that a collection of inanimate chemical compounds, the building blocks of which were cooked in the interior of ancient stars, can come «to know itself». A universe possessing a component that can know itself is, I would argue, more interesting than a universe devoid of such components. Intelligence – however defined and however manifested – is more interesting than inanimate material; sapience is more interesting

[7] M.K. Haba, Y. Lai, J. Wotzlaw, A. Yamaguchi, M. Lugaro, M. Schönbächler, *Precise initial abundance of niobium-92 in the solar system and implications for p-process nucleosynthesis*, in «Proc. Natl Acad. Sci. USA», 118, 8, p. e2017750118, 2021.

[8] C. Sagan, *Cosmos: A Personal Voyage*. Episode 1. PBS, New York 1980.

than sand. And the greater the depth of that intelligence and sapience, and the greater the variety of types of intelligence, the more interesting the universe becomes.

Although it remains an article of faith, particularly when, as philosophers such as Chomsky note, the maintenance of intelligence on planet Earth is far from guaranteed, I argue that in our local corner of the universe intelligence – whether human; human-developed AI; or some merger of human with AI – will come to appreciate the wonder in Sagan's statement that «we are a way for the universe to know itself». At that point, for reasons of altruism rather than selfishness or competition, Dick's Intelligence Principle will come to apply.

I believe the Intelligence Principle as originally stated is, however, missing one vital aspect; it is perhaps implicit in the Principle, but it should be made explicit – particularly if one adopts a Stapledonian mindset. The same astrophysical processes that seeded the primordial cloud from which our solar system developed, and provided the elements from which this specific part of the universe could come to know itself, are equally capable of destroying intelligence, whether biological or postbiological. Supernovae possess the capacity to destroy life on nearby planets; as do gamma-ray bursters; as do magnetars. More local celestial threats come from meteor impact events; cometary impact events; coronal mass ejections.

Even more locally, existential threats for biological life include climate change; pandemics; environmental degredation. Postbiological entities would likely not be immune from local existential threats, albeit those threats would likely be of a different nature. If it is localised to one planetary system, intelligence – whether human; human-developed AI; or some merger of human with AI – is at risk of extinction. I therefore believe the Intelligence Principle requires a small extension: «...to the

extent intelligence can be improved, it will be improved *and it will spread*» (my italics).

This proposed extension resonates with Dick's emphasis on cultural evolution in his original formulation of the Intelligence Principle. The modification reflects the importance of cross-cultural fertilisation and interaction: advancements in intelligence could permeate and influence neighbouring or interconnected societies. (These societies could arise either from independent abiogenesis events, or from divergent evolution of outpost colonies of a single species).

Throughout human history, advancements and innovations have spread across different cultures and societies, influencing and shaping one another; the revised Intelligence Principle extrapolates this to a cosmic scale. The extension also suggests that the improvement of intelligence in one civilisation could trigger a cycle of mutual enhancement across civilisations, leading to a collective and accelerated growth in knowledge in an interstellar setting. The modified principle encourages exploration, collaboration, and discovery; it would foster a culture of shared learning and intellectual advancement among civilisations.

One can of course apply much of the same critique to this proposed modification as one applies to the original Intelligence Principle. The words "and it will spread" assume a deterministic spread of intelligence, which might not always be the case. The dissemination of knowledge and intelligence among different civilisations, certainly at the interstellar level, might be unpredictable and contingent on a number of factors. Furthermore, some civilisations, through choice or circumstance, might remain isolated or unresponsive, thus challenging a universal application of the proposed modification. In addition, as has been long been explored in both academic fora and fictional settings, there are ethical dilemmas involving the uncontrolled spread of intelligence: a civilisation might be negatively impacted by contact with another

civilisation, or at least be influenced, even if inadvertently, in negative ways. Finally, the proposed extension could be seen to oversimplify the challenges involved in spreading intelligence over interstellar distances.

Notwithstanding these points, I argue that, in the absence of any evidence for the existence of intelligence elsewhere, the responsibility of humanity to ensure the universe can continue «to know itself» will lead to our reaching for the stars.

4. *Ethical implications*

In reference to the seeming absence of extraterrestrial civilisations, the great Italian physicist Enrico Fermi once asked: «Where is everybody?»[9] (see Webb 2015 for an overview of this question). Fermi realised that an application of the Copernican Principle – that humans are not privileged observers of the universe; crudely put, that there is nothing special about Earth or our species – suggests there could be a myriad of life-bearing planets in the Milky Way galaxy. Life elsewhere, however, might have started hundreds of millions, perhaps billions, of years before the abiogenesis event here on Earth. If technological life elsewhere followed the path taken by humanity, namely of exponentially increasing technological improvement, and by the Copernican Principle we cannot deny them that opportunity, then Clarke's third law suggests such life could possess abilities 'indistinguishable from magic'. Fermi was hinting that advanced extraterrestrials should already be here; at the very least, when we look up we should see signs of their activity. Yet they are not here; and the universe appears empty, silent.

Researchers have posited several potential drivers for cosmic expansion. Some have argued, for example, that a game-theoretic

[9] S. Webb, *If the Universe is Teeming with Aliens… Where is Everybody?*, Springer, Berlin, 2015.

analysis suggests a less altruistic motive for expansion: the earliest civilisation would attempt to employ first-mover advantage, and populate the galaxy in order to prevent others from doing so. It has also been suggested that sub-elements of society might want to spread a message (a religious idea, a political viewpoint, and so on). Dick's paper refers briefly to the Fermi paradox, and notes that the «roaming of intelligent probes might lead to an AI version» of the paradox, but his main concern is an exploration of what the existence of a postbiological universe might imply for the search for extraterrestrial intelligence (SETI) endeavour.

The proposed modification of the Intelligence Principle to include the compulsion to expand further sharpens the Fermi paradox. If Earth-bound intelligence (human; human-developed AI; or some merger of human with AI) does indeed conclude that expansion through the cosmos is a means of ensuring that the universe can continue to «know itself» – and to know itself in ever-deeper and ever more rewarding ways – then other civilisations would presumably also have reached that conclusion. Those civilisations should already be moving throughout the Milky Way galaxy, increasing the quantity and quality of intelligence in our part of the cosmos. Indeed, the same should have happened in other galaxies: we should see evidence of KII or KIII civilisations. And yet, to date, the universe within a radius of 969 Mpc appears to contain no such advanced technological civilisations[10]

The Fermi paradox admits numerous possible solutions. One straightforward explanation for the absence of observations of extraterrestrial intelligence is that there is nothing to observe: that Earth is currently home to the only intelligence in the galaxy and perhaps the universe. Although the explanation appears to flout

[10] Y. Uno, T. Hashimoto, T. Goto, S.C. Ho, T. Hsu, S. Burns, *Upper limits on transmitter rate of extragalactic civilisations placed by Breakthrough Listen observations*, in «Mon. Notes Roy. Astro. Soc.», 532, 3, 2023, pp. 4649–4653.

the Copernican principle, this does not necessarily follow. The universe perhaps has to possess a certain size in order for observers to exist with the capacity to observe it; that we happen to be the observers would not make us special. If we accept this explanation, however, and in the absence of evidence to the contrary I believe this is prudent, then how should we act?

The modified Intelligence Principle suggests that we should act to spread the ability for the universe «to know itself». In order to do that, however, we must in the first instance *maintain* and *perpetuate* that ability. The Intelligence Principle contains no teleological element; the phenomenon of intelligence, if it is confined to this one planet, could easily be removed from being. The universe would then lose its wonderful ability «to know itself». As Lo Sapio[11] has argued, with our current level of understanding, we should adopt the principle of «survival at any cost» as a guide to our actions.

5. *Conclusion*

As the philosopher David Chalmers argued more than a decade ago[12], humanity may be at the cusp of an "intelligence explosion". Recent technological developments are consistent with Chalmers' observations, and one can imagine the creation of a positive feedback loop in which intelligence makes itself more capable and thus improves its ability to make itself even more intelligent. This scenario does not rely purely upon the existence of an appropriate computational substrate: the augmentation of human intelligence by AI, by genetic engineering, or some combination of the two, might also lead to an intelligence explosion. This event,

[11] L. Lo Sapio, *The ethics of astrobiology: humanity's place in the cosmos and the extinction problem*, in «Frontiers in Astronomy and Space Sciences», doi: 10.3389/fspas.2022.1008265, 2022.

[12] D. Chalmers, *The singularity: a philosophical analysis*, in «J. Consciousness Studies», 17, 9–10, pp. 7–65, 2010.

if it does come to pass, could occur on much shorter timescales than is typically assumed when engaging in Stapledonian thinking.

An intelligence explosion would bring with it attendant dangers. Assuming we successfully navigate the dangers, where might this development lead? This question is of great importance – an intelligence explosion would be one of the most significant events in the history of life on this planet – and yet has not been discussed in academic circles as thoroughly as it deserves, perhaps due to the speculative flavour of the idea.

In this paper I have not addressed the question's many practical and philosophical aspects, and chose instead to focus on one point: that an appreciation by intelligence (whatever form that intelligence eventually takes) of the value in ensuring the universe can continue to know itself will lead to the spreading of intelligence through the cosmos.

This conclusion then encounters the Fermi paradox: would intelligences elsewhere not possess the same drive to spread through our galaxy? And yet we see no evidence for the existence of those intelligences. If we adopt a conservative solution to the Fermi paradox – namely, that we are alone – then a large burden is placed upon us. It becomes incumbent upon humanity to act to overcome the current existential threats facing us, climate change in particular; to successfully navigate an intelligence explosion, ensuring that issues of value and morality are addressed so the outcome is good rather than bad; And then… To reach for the stars.

STEPHEN WEBB is the Head of Technology Enhanced Learning at the University of Portsmouth

stephen.x.webb@port.ac.uk