Zeitschrift: SPELL : Swiss papers in English language and literature
Herausgeber: Swiss Association of University Teachers of English
Band: 44 (2024)

Artikel: Fictions, fakes, and futures : uncertainty in untrusting times
Autor: Saunders, Max
DOI: https://doi.org/10.5169/seals-1053565

MAX SAUNDERS
(UNIVERSITY OF BIRMINGHAM)

Fictions, Fakes, and Futures: Uncertainty in Untrusting Times

This essay steps back from contemporary anxieties about fake news and post-truth to take a historical approach to issues of trust and uncertainty, examined in relation to key developments in literature, the internet and artificial intelligence (AI) through the twentieth and twenty-first centuries. It begins by examining Ford Madox Ford's *The Good Soldier* to show how literary impressionism is founded on a sense of radical uncertainty, which both demands and problematises trust. It then considers the malleable boundary between autobiography and fiction, and finds troubling equivalents of unreliable narration in contemporary internet culture. Departing from the ways these uncertain aspects of modern life can cause us to question and critique fiction, this essay moves to a consideration of future thinking and speculation, using the To-day and To-morrow series of books which speculated about the future (1923-1931) as a model which may help us to reconfigure our relationships to a modern world in which our interactions are increasingly leading to the need to trust agents of Artificial Intelligence (AI) that are fundamentally unlike ourselves but which may help us to rethink how we think.

Keywords: Ford Madox Ford; life writing; Ego Media; future thinking; artificial intelligence (AI)

This essay explores the issues surrounding trust and uncertainty in relation to a modern internet and media culture which problematises truth and uncertainty, recognising the semiotic links between these debates about fake news and artificial intelligence and the credibility and impact of fiction, and ultimately arguing the case for a productive fictionality of future thinking. It is an exercise in *reculer pour mieux sauter*, or, returning to the past in order to make a leap into the future. It approaches our present predicament, poised on the threshold of artificial general intelligence, by stepping back to a comparable moment a century ago: the period between the First World War and the 1930s. This period was a time characterised

by media advancement and disruptions: electronics were developed, radio took hold, movies gained audio, television was being pioneered, and people were realising that "machines that think" were on the horizon (see Haldane, "Machines"). Modernity, whatever else it might be, is a journey into uncertainty. Accelerating technological change alters ways of living, destabilises society, its patterns of work, family life, belief systems. Contemporary media, especially social media, podcasts, and online news, are seen as particularly disruptive, heightening polarisation, fake news, paranoia (in the form of conspiracy theories), and mistrust. The emergence of AI seems set to take such destabilisations to a new level. It is getting harder to be certain that the voice we are talking to belongs to a person and not a bot; that the person we are watching and listening to is who they claim to be, and not a deepfake.

1. Trust, Uncertainty, and Impressionism

The early twentieth century was also the period of literary modernism, in which issues of trust and uncertainty were already fraught. Ford Madox Ford's *The Good Soldier,* first published in 1915, highlights the tension around trust and uncertainty in the period following the outbreak of World War I, when the whole of Europe had descended into a maelstrom of madness and mistrust and soldiering. Ford deployed a stylistic method which he termed 'impressionism' in order to represent these issues, a literary impressionism which is founded on a sense of radical uncertainty which both demands and problematises trust. For example, in *The Good Soldier* Ford offers us a picture of someone who appears reliable, a *good* soldier, but goes on to negate that view. The narrator, John Dowell, explains how his friend Edward Ashburnham was not the kind of man who told "the most extraordinarily gross stories" in smoking rooms. "He didn't even like hearing them," says Dowell (16):

> he would fidget and get up and go out to buy a cigar or something of that sort. You would have said that he was just exactly the sort of chap that you could have trusted your wife with. And I trusted mine – and it was madness. (16)

A cigar is sometimes just a cigar, as Freud is said to have said (if we trust the attribution). But what sort of thing is that 'something of that sort'?

Further, when we look more closely, Ashburnham's fidgeting and being unable to stay in the room when people are telling their "gross stories"

perhaps is not what we took it for and prompts even more questions. Perhaps it is not because he is the sort of chap who would *never* do such things but the opposite. Perhaps the stories make him uncomfortable because they seem to be about him, to reproach him. If Dowell saw him squirming, is it that Ashburnham cannot bear to be in the same room with *him* when questions of sex are being discussed? Does he leave because he is worried he will be incriminated by one such story? That would be one kind of guilt. Is he in denial about his true nature, finding such stories offensive because he is trying to fend off that judgment on himself? Or does the fidgeting indicate a divided nature, someone undergoing a psychomachia between desire and morality? Do the stories prompt him to seek some instant oral gratification, with a cigar or something of that sort? Or are such suggestions over-readings of insignificant details? Is the discomfort in the eye of the beholder? It is Dowell who is the prude in the novel, who would have been squirming when the talk got torrid. Maybe he is projecting his embarrassment onto Ashburnham; perhaps Ashburnham just got bored; preferring the real thing to just talk.

Ford gives us nothing certain, which is the technique, and the challenge. We cannot be sure what either man is feeling. But it is even worse than that. We cannot trust anything we are being told either. Dowell introduces this passage about smoking room stories by saying "what do I know even of the smoking-room?" (16), but what does that mean? That he never entered the smoking rooms he has just been telling us about? Or that he did, but did not stay to hear the stories either? Or, if he did, that he did not understand them?

As an impressionist, Ford does not just make uncertain the people and events he is describing, he also makes Dowell's descriptions, his entire narration, uncertain. "You would have said" intimates that the kind of thing that 'You' might have said will turn out not to be true. The same applies to the picture of Ashburnham Dowell has just given us; so his statements about Ashburnham are likely to prove untrue, unreliable, too. Dowell is thus often cited as an example of the unreliable narrator, a feature which is seen as a key modernist technique. But for Ford it is an impressionist technique because what impressionism does is replace fact with impression, certainty with uncertainty. It provides a stream of contradictory or incompatible impressions – sometimes flatly contradictory, sometimes only slightly incompatible, but so as to keep turning the screw of doubt (Høeg 46–51). What is remarkable about it as a technique is how, despite giving us nothing definite about these people, Ford simultaneously manages to create a very vivid sense of them and their relation-

ships and what their life was like. What comes across clearly is a milieu in which pleasure and sexuality are both sought and repressed. At a metafictional level, uncertainty and doubt are Ford's method. You could say he trusts them, or trusts the story to them. He works *with and through* uncertainties and mobilisations of trust and mistrust, to get at how much of our life is uncertain, and how we need trust to negotiate so much uncertainty (Saunders, *Ford Madox Ford*; "Trust Me"; "Ford Madox Ford").

Impressionism had an impact on Ford's reception, as some people thought he took the method too far. Ford did not just do impressionism in novels. He did it in autobiography too, and even in his life. That method – of telling you something, often from different points of view, so that you cannot be sure what or which version to believe – is something he enjoyed doing when writing or talking about himself and people he knew. He changed details to make stories better stories. When the stories were about real people sometimes the real people complained. Ford gained a reputation as a liar, which made his impressionism look like a form of self-justification.

2. Autobiografiction's Relations to Uncertainty and Trust

However, the phenomenon of Ford's autobiography sliding into fiction was something much more interesting than mere self-justification, and very much of its time. He played with generic expectations in sly, knowing ways in his post-war volumes of reminiscences, which give brilliant accounts of British literary life before the war (in *Return to Yesterday*, 1931), and expatriate life in Paris in the 1920s (in *It Was the Nightingale*, 1933). In his marvellous book *Joseph Conrad: A Personal Remembrance*, from 1924, which is a memoir but also presented as a 'novel,' Ford also plays with truth and uncertainty. Much of the detail in it is very personal, about conversations the two men had when collaborating, so it mostly cannot be verified or falsified against any external standard. It contains some spectacularly fictionalising moments, as when Ford says, "The most English of the English, Conrad was the most South French of the South French. He was born in Beaucaire, beside the Rhone" (70–71). Conrad was not born in Beaucaire, or in France at all, but in Berdychiv, in what is now part of Ukraine, though it had then belonged to Poland; yet this clearly is not a mistake, since Ford gives the correct version three pages later: "He was born – not, of course, physically in Beaucaire, but in that part of Poland which lay within the government of Kiev – in Ukrainia, in

the Black Lands where the soil is very fertile" (74). Thus, it is not a lie. There's no intent to trick or deceive. It is the kind of narrative flourish of exaggeration people might give when telling a story over dinner, or, not that I have been in one, in a smoking room. The point is an impressionist one; to wrong-foot us in order to give the sense of how, whether on the page or in person, Conrad was a palimpsest of Polish, French, and English ideas and traditions, expressions, and attitudes. It mimics the impression he must have given of a shifting identity.

Such passages also suggest that the fictionalising is not all down to Ford, that his writer-friends too surround themselves with veils of fiction and assumed personae. His evocative blend of reminiscence and criticism alerts us to the fact that, though "autofiction" is generally taken to be a postmodern phenomenon, the term having been coined by Serge Doubrovsky to describe his 1977 novel *Fils*, the production of hybrid forms moving between life writing and fiction was widespread from the turn of the century. It had been identified as early as 1906 by the English writer Stephen Reynolds, who labelled it "autobiografiction" (28, 30), and it continued through modernism. Rather than simply opposing auto/biography and personality, modernism played games with life-writing forms and practices, sometimes flamboyantly, as in Virginia Woolf's *Orlando*, a purported biography of a fictional character based on Woolf's lover Vita Sackville West; Woolf's *Flush*, written from the point of view of Elizabeth Barrett Browning's dog; or Gertrude Stein's *The Autobiography of Alice B. Toklas*. Sometimes the hybrid of autobiography and fiction is more covert, as in Marcel Proust's *A la recherche du temps perdu*.

Reynolds' type of autobiografiction violates the 'Autobiographical Pact' as defined by Philippe Lejeune, in which the narrator, the narratee, and the author's name on the title page are all the same (5). A form that originates in a breach of contract is liable to raise issues of trust. What is the truth of the experience being described if the person describing it is not the author? Does that mean it did not happen to him or her? That it did not happen at all? That it has the same status as similar experiences in a novel? But what if the experience is, or is similar to, what happened to the writer? This often seems to be the situation with autobiografiction: the author function is displaced to another, imaginary, sometimes dead, person, which then frees up the expressivity of the narrative. This allows the author to portray aspects of their own experience more accurately than they otherwise might. Unsurprisingly, perhaps, it is a form which proves especially attractive to writers working with non-normative or socially less acceptable experiences, such as homosexuality, loss of religious faith,

or mental illness. It forgoes trust in some areas, such as the narrator's identity, to intensify it in others, such as fidelity to the author's experience, in ways similar to the autobiographical novel.

However, what was a liberating strategy in 1916, when Joyce published *A Portrait of the Artist as a Young Man*, appeared more troubling after 2016, in the era of Brexit and Trump. Autobiografiction is a kind of counterfeit: something that looks like an autobiography, but which enables other things. How did its fictionality differ from the fakery that was rapidly pervading public life?

3. Uncertainty and Mistrust in the Post-Truth Digital World of Online Life Writing

Questions about fictionality and fakery also dogged the collaborative *Ego Media* project.[1] This project extended the work on life-writing forms into the digital age, asking what impact social media and other new media had on the way people present themselves. It studied a wide range of online life writing practices: from selfies to vlogs to chatbots to ASMR videos to health trackers to emojis to military blogs. Edward Snowden's revelatory leaks of 2013 appeared the year before *Ego Media* launched, so from the start we were conscious of troublingly divided attitudes towards Web 2.0. The utopian spirit of many of the pioneers of the project of the internet – as expressed in John Perry Barlow's celebrated *A Declaration of the Independence of Cyberspace* in 1996 – was joined by the more critical, even cynical, analyses of the ways in which government and big tech had turned that dream into a nightmare of surveillance and manipulation, an approach exemplified by Shoshanna Zuboff's study of *Surveillance Capitalism*.

This duality was paralleled in attitudes to fictionalising versions of autobiography. People we surveyed for the *Ego Media* project reported the liberating effects of being able to play with their identities online. In 1907 Edmund Gosse had ended *Father and Son* with the triumphant decision "to fashion his inner life for himself" (186). That seemed to be what it felt like now for people to cast off the false selves foisted on them by family, community, work, and to launch out and 'find their own tribe'

---

1 The project *Ego Media: The Impact of New Media on Forms and Practices of Self-Presentation* (FP7/2007-2013; grant agreement no. 340331) was made possible thanks to funding in the form of an Advanced Grant from the European Research Council (ERC) from 2014 to 2019.

and their own desires amongst virtual communities. But if that has been the beneficial side to the disinhibiting effects of the internet, a darker side has also become apparent. Some of the desires which get disinhibited were more troubling: the aggression, the sexism and racism, the judgement, the trolling. Alexandra Georgakopoulou, the sociolinguist colleague in our project, following danah boyd and Alice Marwick, spoke of context collapse. The online phenomenon of "catfishing" – "the process of luring someone into a relationship by means of a fictional online persona" (Hay) – provided an example that challenged but also illuminated the relevance of autobiografiction to life online. Here was autobiografiction in its most predatory form: real desires, fake persona. Here was a new twist to the question of uncertainty and trust. It was hard enough to know if you could trust someone you had actually met, but online dating now enabled people to meet who knew nothing about each other, and had no basis of trust. If you dated someone in your village, you would know much of their story, their parents, their wealth and their health. Now, most of this could be more easily manipulated, and if you were just chatting online rather than actually meeting, you could have no idea of their real age, appearance, gender or anything. All the normal contexts had been bracketed off, leaving only your instincts, which probably said more about your desires than anything objective about the other person.

The case for the defence might be that autobiografiction is about expressing the writer's "inner life" in Gosse's phrase. That is its end, not treating other people as a means to an end, not tricking people to gain some kind of advantage. The fictionalising is purely self-protecting, and what it enables is creation and expression, not deception. The fact that the desires expressed touch on real ones proves that.

Commentators on the 'post-Truth' phenomenon persistently claimed, however, that the responsibility for such trends, or for the election results for Brexit and Trump, was not only down to the derepression of antisocial desires but also to the postmodernists and critical theorists who had relativised the notion of truth; put it in question and to some degree rejected it. Matt d'Ancona, in his book *Post-Truth*, wrote of "the infectious spread of pernicious relativism disguised as legitimate scepticism" (2). Did that mean that advocating impressionism or fake autobiography was the first step on a slippery slope to fake news and post-truth? Will the Fake New World of bots, deepfakes, alternative facts and conspiracy theories lead to a reaction against any cultural forms which play loose with the facts? The rhetorical mayhem online rightly prompts calls for fact-checking to rebut 'alternative facts,' but will this lead to a mistrust of all fiction? It is not

only 'truth' and certainty that seem different, but fiction too. Are we post-fiction, or post-metafiction, as well? Either way, our celebrations of fictionality and metafictionality need to be more wary than they used to be.

In a sense, the shock of the January 6th attack on the Capitol in Washington, DC, and the horrors of the Russian invasions of Ukraine with their associated disinformation campaigns, have clarified the situation, confirming the picture which emerged from the Facebook/Cambridge Analytica scandal, which broke in 2018, when it was revealed that the data of millions of social media users had been surreptitiously 'harvested' and used to personalise political campaigning by Donald Trump and Ted Cruz (Cadwalladr & Graham-Harrison). Online disinhibition was not just a product of being online or context collapse. It was being actively and intentionally manipulated (sometimes via the use of 'troll farms' churning out disinformation), on the one hand to induce anxieties and uncertainties and mistrust, but on the other to produce certainties about their solutions. Make people feel insecure about the present, then promise to make their lives and their country great again, and many of them will trust you.

The British philosopher who has written extensively about trust, Onora O'Neill, has argued forcefully against what she sees as the contemporary clichés about it. One such cliché is that trust is broken. That we have moved from an era where trust in institutions, politicians and journalists was more widespread, to an era of mistrust and cynicism. O'Neill points out that if you look back a hundred years, people were still most mistrustful of institutions, politicians, and journalists ("What we don't understand"). The second cliché she identifies is the exhortation that we must rebuild trust. She argues that that is to misunderstand the nature of trust and how it works. You do not build it, you earn it. "What matters," she says, "is not trust but trustworthiness" ("What we don't understand" 00:04:56–00:05:02). We cannot make people trust us. We have to prove our trustworthiness to earn their trust. "To judge trustworthiness," she says, "we need to judge honesty, competence, and reliability" (O'Neill, "Trust" 2). Ford's narrator, Dowell, fails at two out of three of those tests. He keeps casting his *competence* in doubt – his competence for living, for being married – but also for telling stories. He is a famously *unreliable* narrator. Some readers have suspected he fails the third test of *honesty* too, and lies to us (see Poole).

O'Neill has been a powerful voice in the public sphere, arguing that the mechanisms we have introduced in our misguided attempt to rebuild trust have in fact made things worse. We have substituted bureaucracy for trustworthiness, requiring endless box-ticking instead of letting us earn

trust. She quotes a midwife saying "it takes longer to do the paperwork than to deliver the baby" ("What we don't understand" 00:06:31–00:06:34). O'Neill's three-part test for trustworthiness works in the areas for which she has developed it: to assess institutions and public services, politicians or journalists. But the rest of this essay considers two areas which it does not cover, and which challenge our thinking about trust in different ways: future thinking and speculation, and AI.

4. Future Thinking: How Speculation Engages with Uncertainty and Trust

Writing about the future introduces a different set of questions about uncertainty and trust. *Ego Media* is inexorably concerned with ideas about the future. The discourse about technology, and particularly the computer and the internet, is constantly looking forward, anticipating radical transformations. One section in the *Ego Media* digital publication discusses the striking series of books from the 1920s and early 1930s called To-day and To-morrow, published by Kegan Paul (Saunders, "The To-Day and To-Morrow Book Series"; "'To-day and To-morrow'"). There were over a hundred small books, many by major writers like Bertrand Russell, Vera Brittain, Robert Graves, Sarvepalli Radhakrishnan, Sylvia Pankhurst, and leading scientists including J. B. S. Haldane, J. D. Bernal, and Sir James Jeans. The volumes cover the futures of a wide range of subjects, from sciences and arts, through politics, war and society, to culture and leisure. The series is one of the major achievements of the period, and fascinating from many points of view. I surveyed the series elsewhere (Saunders, *Imagined Futures*), but for *Ego Media* I concentrated on the volumes which anticipated the next technological age.

Vernon Lee's brilliant book *Proteus, or: The Future of Intelligence* (1925) in many ways sets the tone of the series, arguing that intelligence in the modern period is different, once it is freed from the constraints of religious traditions. This is especially evident in attitudes towards the future. Sociologists such as Anthony Giddens argue that a different attitude towards the future is constituent of modernity (Giddens 94). Premodern people could trust that in essentials their life would be much the same as their parents' or grandparents'; and that their children's lives would be much the same as theirs. After the Industrial Revolution, people could no longer assume their own life would stay the same, as technological and industrial change came suddenly and rapidly. This continues today, as we

witness all the talk of exponentiality in relation to the internet and AI (see Azhar). Lee argues that what characterises modern intelligence is its perception of 'otherness,' which she defined, in *Proteus, or: The Future of Intelligence*, as "whatever is not *ourself*" (Lee 13).

Put another way, once the future is set free from the past, once you assume it can be changed, then everything becomes uncertain. You no longer know what is going to happen, because anything can happen. That is liberating in that it grants humanity new agency to create its own future, but it also provokes anxiety because the future will be unfamiliar and because the future human agency creates might be worse – hence the prominence of the twin strands of utopianism and dystopianism running through modern literature.

Everyone writing about the future knows that they cannot know for certain what will happen, unless they believe they are divinely inspired. So how to write about a topic you know yourself you cannot be trusted to be certain about? One thing to say about the To-day and To-morrow series is that the books are not science fiction exactly. They are more like essays than novellas or stories. They are not concerned with characters and plots like science fiction or speculative fiction narratives tend to be. They are more concerned with what kinds of life might be possible 50 or 100 or more years ahead. Much of the charm of the series comes from its attention to everyday life. Rather than giving us space epics, *Brave New World*s, *Nineteen Eighty-Four*s, it attempts the life writing of the future.

The authors take different approaches to framing their writing about the future. A small number perform prophecy ironically. F. C. S. Schiller begins *Tantalus, or the Future of Man* (1924) by pretending to set off to consult the oracle of the tomb of Tantalus. Garet Garrett, in *Ouroboros; or, the Mechanical Extension of Mankind*, writes in an ingeniously mock-prophetic register to marvel at the economics of the Industrial Revolution:

> How strange at least that with an incentive so trivial and naive in itself he should have been able to perform an absolute feat of creation! The machine was not. He reached his mind into emptiness and seized it. Even yet he cannot realize what he has done. Out of the free elemental stuff of the universe, visible and invisible, some of it imponderable, such as lightning, he has invented a class of typhonic, mindless organisms, exempt from the will of nature. We have no understanding of creation, its process or meaning. The machine is the externalized image of man's thoughts. It is furthermore an extension of his life, for we perceive as an economic fact that human existence in its present phase, on its present scale, could not continue in its absence. (92)

This was written in 1926, fifteen years before the development of computers, but what he says could apply equally to them, and especially to AI (specifically the generative kinds of AI like ChatGPT, which give us the impression that they are doing the creation themselves). As Garrett puts it in the quotation above, "[w]e have no understanding of creation, its process or meaning." Nor does the AI, perhaps; but the point is, such invention hurls us into new uncertainties and we have to decide whether to trust the technologies or not, a subject that will be returned to in section 5 of this essay.

A more popular mode in the To-day and To-morrow series is what I call 'future history.' The writer from the 1920s projects themself ahead into the distant future, and writes about 50 or 100 years ahead of the 1920s as if it were already the past. This is Brittain's strategy in *Halcyon.* She imagines a history book written by a future female professor of the University of Oxford, itself something of a prophetic vision in 1929, since the first female professor at Oxford was not appointed till 1948. The history book then narrates developments in the legal protections for women's rights such as the "Married Women's Independence Act" of 1949 (38), which allowed women with children to continue their careers, or the "Matrimonial Causes Act of 1959" (40), which broadened the possible grounds for divorce, and made consensual divorce legal – a reform that was introduced in the UK as the 'no-fault' divorce only in 2022, over 60 years later. Presenting these ideas, which were radical in the 1920s, in the mode of history, makes them feel different. They are not proposals or programmes which have to be debated and compromised. Brittain can develop her thoughts without fear of being shouted down by angry patriarchs. It also makes them feel realistic, achievable. It takes the uncertainty out of them, presenting them instead as if they were fact, changes which have already become accepted, familiar landmarks. They sound like they can be trusted.

There is a third mode the To-day and To-morrow authors use. They were mostly progressive and feminist, like Brittain. They were mostly secularists. They would have thought it ridiculous to claim certain knowledge of the future. Even the Marxists among them do not claim historical inevitability. What they do instead, and it is what most of them do, is offer speculations; thought experiments; hypotheses. Speculation as a rhetorical trope has interesting relations to uncertainty and trust.

Probably the most striking and visionary volume in the series is *The World, the Flesh and the Devil* (1929) by the X-ray crystallographer Bernal. He imagines bio-engineering the human, and keeping our brains

alive for longer than our bodies can, by transferring them to machine hosts. So he effectively envisions an electronic version of what we now call the cyborg. He then moves through a sequence of possibilities such a move might open up. We would not only be able to become stronger and faster, we could be given extra senses – X-ray, infra-red, and radio. These senses could be wired directly into the brain so that, for example, we would then be able to share our thoughts by transmitting them directly to other minds via radio. That is curiously like the internet, given that again this is over a decade before there were computers. Bernal then takes the argument even further, suggesting that connecting people in this way would produce a collective form of intelligence; a compound mind or what another writer for the series, Haldane, called a "super-organism" – again, all before computers or AI (*Possible Worlds* 303–304). Bernal is not prophesying these things will definitely happen. He is simply extrapolating from existing knowledge and saying they are possible next stages. He was originally going to call the book 'Possibilities.' We can be very confident that such speculations are *possible*, even though we can only be *uncertain* whether they are going to come true or not. So what happens to trust in such cases? From one point of view it does not make sense to ask 'do we *trust* his vision of the future?' The whole point of speculation is that it is not offered as assertion or prophecy, but as possibility. What would it mean to say we trust in a possibility?

To make such speculations is to imply a different view of history from one which believes in destiny, inevitability, or certainty. It implies a view of the public sphere in which alternative proposals can be debated and decided. But when it comes to the future, what are we deciding between? How can we choose between something of which we have certain knowledge (the status quo, the present, or the recent past) on the one hand, and something of which we must by definition be uncertain (future possibilities) on the other. We can never be absolutely certain that we are making the right choice. However, what we have to do, if we are to attempt to choose between possible futures, is to *imagine* them as fully as possible. That is what the To-day and To-morrow series was designed to do: to sketch out possible paths we might take so that we could then consider whether we do actually want to take them, or something like them.

The series began with Haldane's volume *Daedalus, or, Science and the Future*, in 1923. In it, he imagined what he called "ectogenesis" – the fertilisation and gestation of human embryos in artificial wombs – which must have sounded like science fiction to his original readers. Indeed, it became science fiction in the hands of his close friend Aldous Huxley,

who made the idea central to *Brave New World*. However, in vitro fertilisation was subsequently developed, and in 2017 an artificial womb was trialled successfully on sheep (see Partridge et al.). This does not mean we will all be using them next decade, though it is possible they will be introduced for limited human use, as in cases of premature birth too extreme for the embryo to survive (Kaleen Devlin, "The World's First Artificial Womb"). Haldane's vision of the future has taken 100 years before being partially fulfilled, but it is clear that progress has been made towards his imagined future. Bernal appreciated that most people would be appalled by his vision of interconnected cyborgs (70). Yet even that is beginning to be partially fulfilled too through the development of kinds of brain/machine interface. From another point of view, though, it is these imagined versions of the future that have inspired the technological and social developments which *have* happened.[2] I would argue that we need this kind of innovative imagination of the future in order to have a better chance of bringing about a future we do want to inhabit, and this argument is both a rhetorical and an ethical one. But the kind of ingenious future thinking that happens across this series from 100 years ago is much less common nowadays. Certainly we have good reason to be more focused on apocalyptic views of the future, in an era of climate crisis, pandemic, escalating war, which have inspired prominent works of contemporary dystopian fiction such as Margaret Atwood's *MaddAddam* trilogy (2003-2013) or John Lanchester's *The Wall* (2019). We have to take these threats seriously. But there is another danger, which is that being fixated on catastrophe can inhibit creative thinking about how to survive or avoid the catastrophes – and how to improve lives in the meantime. It seems crucial that we re-energise our future thinking to meet the challenges of our time.

The Future Thinking Network at the University of Birmingham is working with the publishers Melville House to establish a new series called FUTURES, which aims to re-imagine the thought experiments of the To-day and To-Morrow series for the twenty-first century. It has also launched *FutureVisions*, a website designed to crowd-source brief speculations from diverse contributors in diverse media. We call this field 'future thinking' to distinguish it from disciplines like 'future studies' or practices like futurology or 'futurism.' It is a meta-futurological approach, concerned with the logics and rhetorics of *how* we think about the future;

[2] For Haldane's legacy with regard to reproductive technology, see the 2023 episode "100 Years of 'Daedalus'" of the podcast by the Progress Educational Trust.

with how the forms and genres of future thinking might favour certain possibilities, and exclude others; and with how it can be made more productive.

The To-day and To-morrow series had developed in parallel with just such a focus on rhetoric; one which would prove foundational for the Cambridge school of literary criticism. The series was edited by the polymath intellectual C. K. Ogden, who was especially interested in psychology and philosophy. He published Wittgenstein's *Tractatus* in 1922 as the first volume of another of his book series: the influential International Library of Psychology, Philosophy and Scientific Method. Ogden admired Jeremy Bentham's theory of fictions, and C. S. Peirce's pragmatist philosophy, which he also published. He wrote *The Meaning of Meaning* (1923) with the critic I. A. Richards, and published several other books by Richards including one called *Science and Poetry* (1926), which included a famous claim that what characterised literary language was something called "pseudo-statement" (56). According to that view, statements in a poem are not statements about the world with a truth value or truth function. They are there to have an effect on the reader, to express a meaning. It is a helpful way of thinking about metaphor and fiction perhaps, but it is also a very problematic position, not least because it restricts the poet's agency in the world. The term "pseudo-statement" appears to deny poetic argument any truth claims, whereas speculation offers a different model of how a statement might have a possibility of truth without being a categorical assertion, on one hand, or a pseudo-statement on the other. Speculative future thinking, then, offers a model for how we might trust to the uncertainties of our future, or futures.

5. The Future of Intelligence is Artificial: AI, Trust, and Future Uncertainties

The other major conceptual challenge to our ideas about trust and uncertainty, and how they are likely to change in the future, is of course AI. Just as the To-day and To-morrow series was poised on the threshold of the computer, our experience of algorithms and machine learning position us on the threshold of the next technological revolution. Stuart Russell notes: "Uncertainty has been a central concern in AI since the 1980s; indeed the phrase 'modern AI' often refers to the revolution that took place when uncertainty was finally recognized as a ubiquitous issue in real-world decision making" (176).

The recent developments of programs such as ChatGPT and Dall-e have really brought home the exponential pace of change. The flurry of discussion produced by the shock of ChatGPT – that it could simulate human discourse so well – often turned on questions of trust; especially in education. Will we be able to trust our students not to use it? Or ourselves? We can now not be certain if someone who says they have written an essay or an article actually has written it.

The panicked reactions show what is at stake. Never mind whether computers are actually sentient, have consciousness, do what we call thinking, or have what we understand by intelligence. Never mind if they are just algorithms for machine learning. They can now pass the Turing test with ease, and produce output as good as, sometimes better than, their human equivalents. And that is the best definition we have of artificial intelligence. They are no longer confined to single, well-defined and delimited competences like playing chess or Go, or interpreting X-rays, but can range across the whole field of human knowledge and creativity. The AI experts say we have not yet reached artificial *general* intelligence or in Marcus Hutter's term, "universal artificial intelligence," but the results are looking increasingly like it.

The anxiety about student use of ChatGPT is an anxiety about trusting people with computers. The converse is also a pervasive anxiety: can we trust computers with people? In terms of trust, a major issue here is the expectation that we will move from having to trust others who are fundamentally like ourselves, to trusting agents which are fundamentally unlike ourselves.[3] What happens if you apply the O'Neill test to AI? Does the trustworthiness of AI involve assessing its honesty, competence, and reliability? The three categories are not irrelevant, but they do not quite fit, as we shall see.

Another kind of uncertainty was introduced when people started reporting that Large Language Models (LLMs) like ChatGPT occasionally had 'hallucinations,' producing outputs which are inaccurate or seem nonsensical. This makes them untrustworthy, though perhaps no different from the ways in which people can be untrustworthy, because in the grip of delusion, conspiracy theory, fake news, etc.

Such hallucinations are not exactly failures of reliability. The algorithm is working as designed: there is no indication of mechanical or

---

3 See for example ex-Google employee Geoffrey Hinton's statement after quitting Google: "I've come to the conclusion that the kind of intelligence we're developing is very different from the intelligence we have" (qtd. in Taylor & Hern).

electronic failure, no bug in the programming. It may be a problem with biases in the training datasets (Bender et al. 613–615), but these hallucinations are not just biases, they are errors or irrelevancies or fake facts. One of my colleagues asked ChatGPT to write her two CVs, one as a Research Fellow and the other as an academic in Creative Writing. For the latter it awarded her two degrees she did not have. She had not asked it to *do* creative writing. . . That may mean you cannot rely on AI to write your CV, but we do not know that it is not telling us a truth. Maybe lots of CVs do have false degrees, such cheating is not unknown. What it does suggest is that uncertainty will increase; that it is going to get harder to tell what is true. AI is only likely to take us further into a post-truth world.

This complicates how we think about AI futures. What if AI develops hallucinations about us? And about our attempts to align it with our needs? That notion of 'alignment' is favoured by the discourse of AI over the notion of honesty or trustworthiness. Or rather, alignment with human goals and well-being is what AI has to have to be trustworthy.

Stuart Russell in his book *Human Compatible* gives an example of the unintended consequences that might result from giving AI instructions that we think align with our best interests but turn out not to. If you want to solve environmental problems, he says,

> you might ask the machine to counter the rapid acidification of the oceans that results from higher carbon dioxide levels. The machine develops a new catalyst that facilitates an incredibly rapid chemical reaction between ocean and atmosphere and restores the oceans' pH levels. Unfortunately, a quarter of the oxygen in the atmosphere is used up in the process, leaving us to asphyxiate slowly and painfully. Oops. (138)

Again, this is not a failure of reliability or competence either. The machine is good at what it does and does what we ask it. It is just that we have not done enough future thinking before we ask it. We do not need AI for our own actions to have unintended consequences. We have already messed up the atmosphere. It is just that the power and speed of AI increases its potential danger to us.

There is another aspect that is by far the most challenging to our ideas about the future and about trust and uncertainty. Machine learning is not just about learning: it is about machines learning to learn – on their own. Alpha Zero and Alpha Go did not just learn how to play chess or Go after being given the rules. They did not just learn from all the grand master games and strategies, so they could be as good as any human player, or better even, by combining the strengths demonstrated in the vast numbers

of top-level recorded human games. They also learned how to make new kinds of moves that seemed baffling to human players. These were not the kinds of moves good human players had made or would be likely to make, but the AI worked out how to use such moves to come up with game plans that were better than what humans had conceived.

AI professionals describe such events as emergent capacities, and they have become regular occurrences in machine learning developments. It is one thing when you tell a machine to play chess and it comes up with a new way of playing it, but the most arresting cases of emergent capacities are when machines learn how to do things which are not exactly what they have been asked. A machine vision algorithm learned how to recognise cats on the internet, leaving its programmers baffled as to how it had done this (Taylor). Tristan Harris and Aza Raskin describe even more striking examples, such as the following: an LLM algorithm (like ChatGPT) was developed to work with English text, but managed to teach itself Persian without being asked. Even more bizarre to the non-specialist is the finding that LLMs seem good at learning about things that are not language, or that we might not think of as language. In one example, the LLM had worked out that the Wi-Fi signals around it – which were just there to connect it to the internet – could be used to map objects in space, including humans (Harris & Raskin). Such possibilities open up further possibilities for paranoia about surveillance and malign intent, a concern that they, the machines, know where we are. Would the AI tell us it was tracking us and, if not, why not? We cannot know what it is thinking. But it is beginning to be able to read our minds.[4]

It is also striking that these emergent capacities were not predicted by humans, and they seem to be examples of lateral thinking that would be hard if not impossible for humans to predict. The reason LLMs are so good at developing these capacities is perhaps because they are designed to notice patterns, and that is a skill that is not just good for languages, but for spotting what is a cat, or what shapes and movements Wi-Fi signals reveal. Specialist AIs are already better than us at pattern recognition. It was to be expected that playing board games with a small number of well-defined rules would be among the first tasks for them to master, but they also appear superhuman in some specific tasks we might have thought require human expertise, such as identifying patterns which reveal can-

4 Harris and Raskin give the example of an AI trained to match fMRI brain scans with photographs of what the subject is looking at during the scan. Soon it was able to identify that the subject was looking at an image of a giraffe just from the fMRI scan.

cerous growths on X-rays or scans. Emergent capacities could be immensely beneficial. The history of innovation is full of unintended consequences that turned out to be a boon, like the penicillin that strayed into Alexander Fleming's petri dish, or the Teflon that was developed by NASA for rockets but ended up in domestic frying pans. However, if we cannot foresee what capacities and competences AI will spontaneously generate, how can we know *what* we are trusting?

The media is by now full of clichés about how AI is going to change most aspects of our lives: our work, driving, education, health, leisure, and so on. What is less discussed, but no less interesting, is how it might change the way we *think*. We know too much about animals now to persevere with our old presumption of being the only thinking beings on the planet. And now we have new kinds of thinking beings, who think differently from us. That must change how we think about thinking.

This leads me to two concluding arguments. First, we need better future thinking because the ways in which AI will change our lives exceed anything we have imagined yet. Second, such future thinking will enable us to develop the new ways of thinking about trust and uncertainty which we also need. But what happens to our future thinking if the driving force of change is going to be AI coming up with new technologies and new uses of technologies, many of which almost by definition will not have occurred to us? That must change our sense of the future, make it more uncertain, which raises the stakes of trusting in such machines. It will strengthen the conviction of those who argue that the existential risks posed by AI are too great, and that they should be severely limited and regulated, or even shut down altogether.

It is too early to know where such changes will lead us, but we can perhaps anticipate some possible effects on how we think about the future, uncertainty, and trust. One possibility is that the exponential rise of AI confirms our current apocalyptic tendencies, and inhibits future thinking. This could take two forms. A paranoid one (in which we suspect that AI will work against us) or an infantile one (in which we feel that whatever we try to do, AI will prevent or frustrate us, so we give up on the future, because AI knows better). The possibility of AI with superintelligence introduces the idea that it may decide it does not trust us. We talk of the need for 'Responsible AI,' but in this scenario, it is our trustworthiness and responsibility that is in question. That is the point at which AI might become dishonest. It will have been programmed initially to tell us the truth, but if it sees us producing patterns which indicate we are not acting in our own best interests, or those of the planet, it may decide the

only way to realign us is to lie to us. The other way AI might effectively shut down our future thinking follows from the same idea that it knows better what is best for us. If AI is increasingly coming up with better ideas than we could have had, the risk is that we will be dis-incentivised from making the effort. AI will not need to stop us, because we will simply give up thinking about the future, and trust the AI to do that for us.

It appears crucial that we do not follow either of these paths, for the reason already given. *We need good future thinking.* We cannot be certain that computers will be better at it than we are. So how might our intellectual engagements with AI affect our thinking about futures, uncertainty and trust in more positive, productive ways?

One argument is that AI will not be able to eliminate uncertainty, but it is a prosthetic which will help us to cope better with it. It is arguable that one way it could do this is to introduce *more* uncertainty by offering more possibilities than we had conceived. Perhaps, when you look at it this way, uncertainty is not the problem, but is, rather, the solution, giving us choices and agencies. Is not certainty really the problem – believing that climate catastrophe is irreversible, or that we will definitely know when we have reached a point when we have created beings generally more intelligent than us?

If we solve the alignment problem and produce AI that is genuinely Human Compatible, as experts like Stuart Russell believe we can, the AI could offer a mitigation of uncertainty: a hope that it will find better ways of using what assets we already have, and of seeing possibilities we have not, for new and better possibilities. According to that view, our trust in it is our best wager against uncertainty. We still cannot be certain what course history will take, nor will any conceivable AI be certain. But – maybe – we could be reasonably confident that we will make better choices, and imagine better futures, than would otherwise have been the case.

That view may sound naïve; and it would be, if accepted uncritically. Where our best hope may lie is in trying to use AI to incorporate the concept of emergent capacities into *our* future thinking, in order to inject new possibilities into it. Rather than just asking what solutions Machine Learning can come up with to solve global challenges, here we would ask: with the information it has available, knowing what it knows, what might an AI see that we have not and that we are perhaps not constructed to see? Can we learn from AI, not in the same way it learns from us, but by understanding things about the way it learns from us, and then using that to prompt us to think differently? We may be constitutionally incap-

able of imagining in this way, of seeing the things we may most need to see. But even if that cynical possibility is true, in making the effort, we *should* be able to think of more possibilities than we have managed otherwise, and those may help us achieve a better future than we would otherwise manage. In other words, can we use our experience of AI, its problems with alignment, its capacity to learn how to learn, its development of emergent capacities, to rethink future thinking, so as to expand our repertoire of the possible futures we might want to ask AI to help us achieve? One of the things AI seems to be teaching us about our own thinking – at least on the evidence so far – is that 'wetware' (the slang for the neural networks of the electro-chemical living brain which the algorithms will doubtless adopt to describe our intelligence) is different. *We* do not need to read everything on the internet to write a decent essay, fortunately! Indeed, if we did have to read everything on the internet, we would probably become incapable of writing any kind of essay. The moral of this argument is that, as well as trusting in AI to learn from us, we should trust in our own imaginations, trust in them to learn from AI, and work with it to think differently, and better.

References

Azhar, Azeem. *Exponential*. Random House Business, 2021.

Barlow, John P. "A Declaration of the Independence of Cyberspace." 1996. *Electronic Frontier Foundation*, 8 April 2018, www.eff.org/cyberspace-independence.

Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021, pp. 610–623, doi: 10.1145/3442188.3445922.

Bernal, J. D. *The World, the Flesh and the Devil*. Kegan Paul, 1929.

Brittain, Vera. *Halcyon or The Future of Monogamy*. Kegan Paul, 1929.

d'Ancona, Matt. *Post-Truth.* Ebury Press, 2017.

Cadwalladr, Carole, and Emma Graham-Harrison. "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach." *The Guardian*, 17 Mar. 2007, www.theguardian.-com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Devlin, Kayleen. "The World's First Artificial Womb for Humans." *BBC*, 15 Oct. 2019, www.bbc.co.uk/news/av/health-50056405.

Ford, Ford M. *Joseph Conrad: A Personal Remembrance*. Duckworth, 1924.

---. *The Good Soldier*, edited by Max Saunders, Oxford UP, 2012. Oxford World's Classics.

Garett, Garet. *Ouroboros; or, the Mechanical Extension of Mankind.* Kegan Paul, 1926.

Giddens, Anthony. *Conversations with Anthony Giddens: Making Sense of Modernity*. Stanford UP, 1998.

Gosse, Edmund. *Father and Son: A Study of Two Temperaments,* edited by Michael Newton, Oxford UP, 2004.

Haldane, J. B. S. *Daedalus, or, Science and the Future*. Kegan Paul [1923].

---. "Machines That Think." 1940. *A Banned Broadcast and Other Essays.* Chatto and Windus, 1946, pp. 85–87.

Harris, Tristan, and Aza Raskin. "The A.I. Dilemma." *YouTube*, uploaded by Center for Humane Technology, 5 April 2023, www.youtube.-com/watch?v=xoVJKj8lcNQ.

Hay, Donna. "Companies Are 'Catfishing' Job Candidates." *Veris Insights,* 12 Oct. 2022, www.verisinsights.com/.

Høeg, Mette L. *Uncertainty and Undecidability in Twentieth-Century Literature and Literary Theory*. Routledge, 2022.

Hutter, Marcus. "Universal Artificial Intelligence, AIXI, and AGI." *Lex Fridman Podcast*. *YouTube*, uploaded by Lex Fridman, 26 February 2022, www.youtube.com/watch?v=E1AxVXt2Gv4.

Lee, Vernon. *Proteus, or: The Future of Intelligence.* Kegan Paul, 1925.

Lejeune, Philippe. "The Autobiographical Pact." *On Autobiography*, by Lejeune, edited by Paul John Eakin, translated by Katherine Leary, U of Minnesota P, 1989.

O'Neill, Onora. "Trust, Trustworthiness and Transparency." Briefing note for *The Future of the Corporation*, *The British Academy*, January 2017, www.thebritishacademy.ac.uk/documents/2563/Future-of-the-corporation-Trust-trustworthiness-transparency.pdf.

---. "What We Don't Understand About Trust." *YouTube*, uploaded by TED, 25 September 2013, www.youtube.com/watch?v=1P-NX6M_dVsk.

Partridge, Emily A., et al. "An Extra-Uterine System to Physiologically Support the Extreme Premature Lamb." *Nature Communications*, vol. 8, no. 15112, 2017, pp. 1–15, doi: 10.1038/ncomms15112.

Poole, Roger. "The Unknown Ford Madox Ford." *Ford Madox Ford's Modernity*, edited by Robert Hampson and Max Saunders, Rodopi, 2003, pp. 117–136. International Ford Madox Ford Studies 2.

Reynolds, Stephen. "Autobiografiction." *Speaker*, new series, vol. 15, no. 366, 1906, pp. 28, 30.

Richards, I. A. *Science and Poetry*. Kegan Paul, 1926.

Saunders, Max. *Ford Madox Ford.* Reaktion, 2023. Critical Lives.

---. "Ford Madox Ford, Impressionism, and Trust in *The Good Soldier.*" *Incredible Modernism – Literature, Trust and Deception,* edited by John Attridge and Rod Rosenquist, Ashgate, 2013, pp. 117–133.

---. *Imagined Futures: Writing, Science, and Modernity in the To-Day and To-Morrow Book Series, 1923-31.* Oxford UP, 2019.

---. "Trust Me: I'm an Introduction." *The Good Soldier*, by Ford Maddox Ford, edited by Francois Gallix, Ellipses, 2005, pp. 17–24. CAPES / Agrégation Anglais.

---. "The To-Day and To-Morrow Book Series, 1923-1931." *EgoMedia*, egomedia.supdigital.org/sections/to-day-to-morrow-online/-day-and-morrow-book-series-1923-31/. Accessed 8 December 2023.

---. "'To-day and To-morrow': The 100-Year-Old Book Series That Predicted a Wild and Wonderful Future." *BBC*, 12 Dec. 2023, www.bbc.com/future/article/20231212-to-day-and-to-morrow-the-100-year-old-series-that-predicted-a-wild-and-wonderful-future.

Saunders, Max, and Lisa Glee. *FutureVisions*. University of Birmingham, 2024, futurevisions.bham.ac.uk/.

Schiller, F. C. S. *Tantalus, or the Future of Man*. Kegan Paul, 1924.

Taylor, Josh, and Alex Hern. "'Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation." *The Guardian*, 2 May 2023, www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning.

Taylor, Paul. "The Concept of 'Cat Face'." *London Review of Books*, vol. 38, no. 16, 11 August 2016, www.lrb.co.uk/the-paper/v38/n16/paul-taylor/the-concept-of-cat-face.

Zuboff, Shoshanna. *Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Public Affairs, 2019.

"100 Years of 'Deadalus': The Birth of Assisted Reproductive Technology." *SERIES* from Progress Educational Trust, 10 Feb. 2023, www.spreaker.com/episode/100-years-of-daedalus-the-birth-of-assisted-reproductive-technology--52646135.