Utopia, Artificial Intelligence, and the Future of Justice

Miles Brundage
May 3, 2015
JUS 691: Futures of Justice
Theoretical Perspectives on Utopia
Revised Critique

Utopia, Artificial Intelligence, and the Future of Justice

Contemporary discussions of the future of artificial intelligence (AI) and robotics often involve optimistic visions of the possible future of human society. For example, in an open letter entitled "Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter" written and signed by leading experts in the field of AI, the signatories endorsed the following statement:

> "There is now a broad consensus that AI research is progressing steadily, and that its impact on society is likely to increase. The potential benefits are huge, since everything that civilization has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools AI may provide, but the eradication of disease and poverty are not unfathomable. Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls." (Future of Life Institute 2015)

What does it mean, however, to say that the eradication of poverty is "not unfathomable" and how does it relate to utopianism? Is it in fact a well-conceived utopian ideal worth supporting, or is utopianism of any sort such as this a danger that should be avoided? In this essay, I explore utopianism, work, and the future of justice from a variety of perspectives, seeking insight into these question and others. First, I will explore the purported dangers of utopianism, which will require critically engaging with the theoretical and empirical literature on utopianism, as well as examples from science fiction that shed light on the nature of utopia, justice, and their possible futures. Through this analysis, I find that utopia as a political project is sufficiently flexible and heterogeneous to not be dismissed as necessarily tyrannical, simplistic, or otherwise intrinsically problematic, motivating its further exploration. Second, through an exploration of Ahmed's *Willful Subjects* and Le Guin's *The Dispossessed*, I will explore the complexity of justice's attainment. This analysis will identify the need for conceiving of utopia in terms of processes as well as outcomes, and thinking through the relationship between utopian goals and utopian subjectivities. Finally, with this foundation established, I will return to the question of the future of AI in relation to the future of society. In this analysis, I will show that insufficient attention has been paid to the complexities of work and justice, as well as the relationship between processes (in particular of subject formation) and outcomes, in popular visions of a "post-work" future enabled by AI and robotics.

What is Utopia(nism)? A Preliminary Exploration

To begin a serious discussion of utopia and utopianism, it's essential to have a theoretical framework for doing so,

which might seem to be aided by a crisp definition of these terms. However, the crispness of such definitions is itself contested and problematic—that is, some critiques of utopianism (such as that found in Kitch 2000, to be explored later in this essay) posit core components of utopianism that purportedly lend themselves to unjust outcomes when utopian goals are realized. Additionally, as Mannheim (1936) reminds us, utopias have taken many forms, and he notably defines utopia in relation to ideology, which also is not fixed in space or time.

Preliminarily for the purpose of the present discussion (while noting that the concept needs to be elaborated and contextualized appropriately in space and time), I define utopia as a space in which justice has been realized (or a vision for such a space), and utopianism as the system of belief that valorizes one or more utopias as worthwhile political projects. Below, I will elaborate on some aspects of this conceptualization through a critical engagement with Kitch's (2000) opposing definition of utopianism, but for now, two points should be noted: first, this definition is ambiguous as to where/when the space in question is located, to allow for the possibility of present and past utopias. The etymologically derived sense of utopia as being by definition beyond space and time begs some of the questions that a substantive engagement with the concept is needed to resolve (such as the realizability of utopian visions). Second, my definition does not explicitly specify that in a utopian space, justice has been realized to the fullest possible extent. If a conception of justice affords degrees of realization, then it can be sensible to say that in a particular utopian space, justice has been realized, without that society being perfect in all respects. If perfection is by definition a component of utopia qua utopia, then certain critiques of utopianism (such as those explored in the section "Kitch and the Maligning of Utopianism" below) become more valid, and nuanced advocacy of utopian aims becomes a priori discredited. This point is emphasized by Gordon (2004), who notes, "in order to even approach utopian thinking we have to stop associating the utopian with the impossibly idealistic (or with its evil materialist twin, the 'never enough'), and using it as a weapon against others and ourselves." (p. 122). Thus, at this point in the analysis, I bracket the question of perfection and a-temporality/a-spatiality as defining features of utopia(nism), and focus more on the specific justice and subjectivity issues at stake in striving towards a much more just (of still imperfect) world. To do this, the works of Gordon, Bloch, and Mannheim are helpful in mapping some of the political issues at stake in contestations of utopia and utopianism.

Gordon, Bloch, Mannheim, and Utopian Politics

Avery Gordon analyzes the challenges facing advocates of a better world in light of the fact that "[t]he utopian has

a bad reputation in many circles, especially radical intellectual ones." (Gordon 2004, p. 113). This reputation, she finds, is less about utopia qua utopia than "mere utopia." This term is found in Herbert Marcuse's work, which states, "When truth cannot be realized within the established order, it always appears to the latter as mere utopia," (quoted in Gordon 2004, p. 121). To Gordon, the bad reputation of utopia can be partially accounted for as a result of the common conflation of "utopia" and "mere utopia," with the latter being "nowhere we can really live, the impossible, the unrealizable future, unrealistic dreams, a luxury for those who can afford to be impractical, 'a breeder of illusions and therefore, inevitably, of delusions." (Ibid). Further, utopian visions are often assumed (even defined) by utopianism's critics as necessarily destructive if realized–hence, if such a conceptualization were true, vigilance against utopianism is warranted. Gordon argues, "That it is our self-appointed duty to protect the future from our 'deformed and repressed social habits' is more accurately viewed, I think, as the conceit which legitimates the deferral and disqualification of the utopian." (Gordon 2004, p. 127). While not resolving the aforementioned definitional issues regarding utopia and utopianism, Gordon's analysis points to part of what is at stake in the present discussion—the legitimacy of certain types of ideas and practices. To the extent that utopia is defined as inherently flawed and dangerous, utopianism's critics stand on more solid ground, and conversely, a conceptualization of utopia that eschews conflation with "mere utopia" is more defensible—hence her insistence on reclaiming the term. To further enrich our understanding of utopia and utopianism and their political stakes, we can draw on the works of Ernst Bloch and Karl Mannheim.

Bloch (1954) articulates a rich conceptualization of utopia in the context of Marxism, art, philosophy, and other relevant literatures and practices. Central to his theory is a nuanced account of the Real and of possibility, which in turn can be disaggregated into a number of narrower terms/frameworks (e.g. the "factually-objectively possible," the "formally possible," and so forth). Bloch connects macroscopic conceptualizations of and projects working toward utopia (e.g. toward socialism) with individual-level utopian longings (e.g. daydreams)—he sees the former as grounded in and legitimized by the latter, in the following sense: humans, unlike other animals, live in a world of possibility and of choice. It is this constant grappling with possibility that defines humanity to Bloch. He writes, "Man is that which still has much before it. He is repeatedly transformed in his work and by it. He repeatedly stands ahead on frontiers which are no longer such because he perceives them, he ventures beyond them." (Bloch 1954, p. 246). This identification of the human essence with possibility and change, however, comes at a cost, namely that there is always a possibility of negative change. This helps explain the hesitations with and outright rejections of utopianism, if the latter is associated with significant change to the existing order. Possibility can become Nothing, in Bloch's terms, if not for the "brave" choices we must make:

"But since in man active capacity particularly belongs to possibility, the display of this activity and bravery, as soon as and in so far as it takes place, tips the balance in favour of hope. Bravery in this sense is the counter-move to the negative possibility of the wrong path into nothingness. But it is only a counter-move that, unlike the rash, abstract heroic deed, it secures for itself the most precise mediation with the given conditions. That is: mediates itself with the ripeness of these conditions and with their content which is on the social agenda." (Bloch 1954, p. 247).

Several aspects of this passage from Bloch are noteworthy with respect to interrogating the politics of utopia. First, there is the idea of hope. Hope, which Bloch links with utopia at individual and social levels, is arguably fundamental to the human condition. Hope for a more just world in particular is a common feature of human history. Cedric Robinson's analysis of the history of socialism finds a "socialist impulse" rooted in the "persistence of the human spirit" that "domination and oppression inspire...in ways we may never fully understand." (Robinson 2001, p. 157). If this is the case, then the a priori rejection of utopian visions runs counter to fundamental human longings, and is thus politically problematic if not delusional, contra the radical intellectuals Gordon references. Second, the bravery Bloch references, while only partially explained, is linked in his analysis to "mediation with the given conditions" which can be more or less "ripe." In this regard, the specific nature of utopian goals, and their (un)realizability, cannot be decided definitionally but instead ought to be grounded in the actual conditions of society, a task I take up in the context of AI later in this essay. Finally, insofar as there is always a risk of nothingness due to the pervasiveness of possibility and choice in the human experience, Bloch's analysis points toward the need for critically examining subject formation viz-a-viz utopia. That is, if bravery (or, as will be discussed in the context of Ahmed's work later, willfulness) is a necessary component of positive social change, then an accurate analysis of utopianism and its (dis)contents requires a consideration of utopian subjects.

Before turning to Kitch's critique of utopianism, Mannheim's analysis of ideology and utopia can help us better understand the nature of utopianism and its spatial/temporal dimensions. To Mannheim, ideology and utopia are intimately related, in that the former constitutes the dominant conceptual framework and the latter is counterposed to a given ideology, identifying its limitations. He writes:

"What in a given case appears as utopian, and what as ideological, is dependent, essentially, on the stage and degree of reality to which one applies this standard. It is clear that those social strata which represent the prevailing social and intellectual order will experience as reality that structure of relationships of which they are the bearers, while the groups driven into opposition to the present order will be oriented towards the first stirrings of the social

order for which they are striving and which is being realized through them. The representatives of a given order

will label as utopian all conceptions of existence which *from their point of view* can never be realized."

(Mannheim, 1932, p. 164, original emphasis).

Here, Mannheim enriches our understanding of utopia and utopianism by grounding it in specific spatial/temporal contexts.

As suggested in my definitional analysis earlier, to define utopias as necessarily outside of space and time and therefore

impossible prejudges their realizability, making more nuanced analysis of utopian projects impossible. In contrast,

Mannheim aids analysis by locating the (purported) unrealizability of utopias more specifically in the dominant groups in

society. Through a historical and sociological analysis of Chiliastic, liberal-humanitarian, conservative, and socialist-

Communist utopias, Mannheim finds that utopias and ideologies have changed substantially over time. As ideologies

change, new utopias are necessary to challenge the dominant social strata and their limited and limiting ideas regarding

what's possible. Utopian ideas not only transcend the here and now (as do many ideas) but more specifically, they also "tend

to shatter, either partially or wholly, the order of things prevailing at the time." (Mannheim 1932, p. 192). This characteristic

of utopianism points to both its central importance and one reason underlying the bad reputation Gordon alludes to.

However, to more deeply understand the challenges facing utopian thought and practice, we now turn to Kitch's critique of

utopianism as an intrinsically problematic way of thought.

Kitch and the Maligning of (American) Utopianism

In *Higher Ground: From Utopianism to Realism in American Feminist Thought and Practice* (Kitch 2000), Sally

Kitch analyzes utopian visions from the perspective of feminism and its history. Seeking and finding a core of American

utopianism that she finds problematic, Kitch sets out to improve the prospects for feminism by moving it towards a post-

utopian orientation. To better understand this critique, we begin with her terminology: "*utopianism*—the mind-sets,

attitudes, assumptions, and goals associated with utopian social-change strategies...*utopists* [are] those who start or

participate in utopian experiments...*utopias* [are] the actual or fictional communities established on utopian principles."

(Kitch 2000, p. 22). This framework maps well onto and complements the earlier definitions introduced in this essay,

though is notably silent on the precise nature of the referenced mindsets, participants, and communities, which she

subsequently elaborates. Building on this schematic, Kitch develops a sustained analysis of many utopias and utopists,

finding common threads that unite them—the aforementioned "mind-sets, attitudes, assumptions, and goals." Notably,

Kitch's analysis, like the one presented here, eschews a formal definition of what precisely "utopian" social-change strategies inherently consist of, focusing her critique on the overarching mentality that she finds unifies them. In part, Kitch sees the danger of utopianism in the hope it purportedly gives rise to, and cannot subsequently realize: "America promises utopia, and its failure to realize that promise makes its sins seem much worse than bigger sins seem in other places. Where the promise is greatest, so too is the potential for disillusionment." (Kitch 2000, p. 27) Like the conflict between utopias and ideologies, this insight helps us understand the critique of utopia counterposed to e.g. Bloch's hopeful characterization of it. If utopianism inherently falls short in its aims and/or has aims that necessarily lend themselves to that failure, then Kitch's point here is highly critical. However, if, as I will soon suggest after reviewing her key points, she overgeneralizes from particular instances of utopias and utopists in her construction of a theory of utopianism, then Bloch's, Mannheim's, and Gordon's articulations of utopia(nism) retain their merit.

Like Mannheim's analysis of utopia, Kitch's construction of a core of utopianism is rooted in historical examples, though in the latter cases ones drawn specifically from American experiences with utopias. Consistent with her focus on the relationship between utopianism and feminism, Kitch particularly attends to utopian communities that had ambitious stated goals related to gender equity. For example, utopian communities such as Oneida failed in their stated goals, often in ways that were not merely unproductive but actively counterproductive with respect to those goals. She writes, "Gender reforms in some communities, such as Oneida, actually increased women's subordination to men because female empowerment was equated solely with liberal sexual practices that men typically controlled." (Kitch 2000, p. 36). Through these examples, Kitch catalogues a variety of failings in utopian communities resulting from, e.g. excessive focus on present problems, which, when utopian visions had a chance to be realized, led to counterproductive results. In reviewing these failings, Kitch summarizes some she considers central to utopianism which have harmful consequences for feminism in particular, namely "its tendency to concretize ideas, to transform ambiguity and contingency into absolute...becom[ing] fixed and [remaining] unexamined, even if utopists themselves are diligently self-critical about their own adherence to those founding principles. … its present focus....[its] appeal to self-interest....[and] its particular approaches to the acquisition and validation of knowledge." (Kitch 2000, p. 50-52). Juxtaposed with these failings is Kitch's alternative, which she calls realism, which "leads us to disavow the discourse of perfection and attend to the task of justifying feminist knowledge claims. It encourages us to consider balancing apparently oppositional concerns rather than casting our lot too hastily with one or the other." (Kitch 2000, p. 56).

In Kitch's work, we find a serious and enlightening critique of utopian thought and practice, particularly in America

and with respect to its relationship to feminism. However, in the context of our present analysis of utopianism in general, two limitations of this analysis need to be noted. First, while useful, critiques and analyses of utopianism based on historical experiences like Mannheim's and Kitch's are importantly limited in what they can tell us about the future. Indeed, if utopias are (at least in part) developed in response to prevailing ideologies, then the specific content and form of utopias and utopianism cannot be decisively critiqued on the basis of utopias formed in past social conditions, or even present ones. This limitation of Kitch's critique does not imply that present or future utopias will avoid the characteristics she finds, but they call into question its broader relevance.

Second, it is not obvious that the characteristics of utopianism and realism Kitch lays out are mutually exclusive or that the latter constitute a natural kind sufficiently coherent to merit wholesale critique (or, indeed, acceptance). If, for example, the prevailing ideology of society were overly concretized, un-self critical, present-oriented, dependent on self-interest, and so forth, a utopia founded against that ideology and emphasizing the opposite of those characteristics needn't meet Kitch's critique. Indeed, utopian visions vary in many important ways that bear on the generality of Kitch's critique. It may be the case that, pace Kitch, "non-plans are still plans," but not all plans are as detailed as others, and there may be many dimensions of variation along which we can sensibly distinguish plans and utopian visions, such as the extent to which they allow for diversity and choice in the implementation and ongoing reconceptualization of the ideology of the society in question.

In light of the foregoing defense of utopianism in at least some of its possible variations, we now turn to Ahmed's analysis of willfulness and science fiction examples. These will help to illustrate the changing nature of utopias that are appropriate to different times, spaces, and subjects, and provide a foundation for the application of these concepts to the case of AI and the future of work.

Ahmed, *The* Dispossessed, and the Processes of Justice

Ahmed's *Willful Subjects* (2014) is an incisive analysis of the question of willfulness—its meaning, history, and political import. Beginning with a Grimm tale about a willful child, Ahmed traces the history of the idea of willfulness and the uses to which it has been put. She explains the role played by willfulness in "poisonous pedagogy" aimed at breaking the will of children and making them into docile subjects, noting that this parallels efforts in other spaces and at other scales of society that are oriented towards instilling willing subjects. The connection between willingness and willfulness are

crucial here in thinking through the future of justice and the meaning of utopia. Ahmed argues that willful subjects, such as feminist killjoys and the aforementioned willful child, are labeled and treated as such on the basis of their unwillingness to conform to social expectations of them. This connection between the willful child and willful adult subjects is an obstacle to justice, and thus merits our attention here in an examination of utopianism. Ahmed writes, referencing an example of Kant's derogatory remarks regarding working classes, "The civilized and educated subjects remove themselves from the very signs of willfulness, from the capricious and the impulsive, as a way of distancing themselves from the lower ranks, from those who are not European, not bourgeoisie, and not male. The less civilized adults (working class, racial others, women, and of course some embody more than one category of less) are thus figured not only as childlike but *as* willful children." (Ahmed 2014, p. 94, original emphasis). Ahmed details examples of willfulness throughout history and in fiction, which together constitute what she calls a "willfulness archive." This idea complements e.g. Mannheim and Kitch's analysis by emphasizing the need for historical perspective in realizing justice. The ideas and practices of the past, the present, as well as those in fiction, can be used as sources of inspiration and self-criticism to find the flaws in utopian projects and visions— if not eliminating those flaws entirely, then at least moving forward with a more clear sense of the challenges and affordances for social change.

In response to the history of the uses of willfulness as a concept for subjugating others Ahmed calls for the term's reclamation, and mounts a case for the political importance of willfulness. The importance of this theoretical and political concern with and focus on willfulness derives in part from its relationship to inequality, making it especially relevant for our discussion of utopia and justice. As Ahmed notes, "*You do not need to become self-willed if your will is already accomplished by the general will*." (Ahmed 2014, p. 150). This insight is especially relevant as we turn to Ursula Le Guin's *The Dispossessed*, which focuses on the experiences of a subject (Shevek) whose will, unlike others on his planet of Anarres, is not accomplished by the general will. More generally, this exploration of *The Dispossessed* will illustrate the complexity of justice's realization (and, following the definitions used here, the attainment of utopia) and the need for constant vigilance against new forms of injustice. In the terms of the earlier discussion, this vigilance can come from the critiquing of ideology with new utopian visions a la Mannheim, a rigorous attention to the flaws in one's and one's group's theories a la Kitch, and a cultivation of willfulness a la Ahmed.

In *The Dispossessed* (Le Guin 1974), the planet of Anarres represents the apparent realization of many of the goals of contemporary utopists. The planet was colonized by Odonian anarchists several generations prior to the beginning of the events of the novel, allowing time for a highly complex social system to have developed, of which Le Guin explores both

the virtues and limitations. By depicting the main character Shevek's life on Anarres and his journey to the more capitalist and socially stratified planet of Urras, from which the Odonian anarchists originally came, we see that the inhabitants of those planets have starkly different attitudes towards work and justice. The Anarresi, for example, use the same word for work and play. Having grown up in this linguistic and conceptual context, the Anaressi are more attuned to the benefits of their society, and more skeptical of the Urrasi's capitalist society, where work is conceived and lived by subjects as burdensome and oppressive (in the lower class, that is—the upper classes, living on accumulated wealth, have no such urgent need to work to survive). When Shevek sees productive workers on Urras, Le Guin writes of his reaction, "He had assumed that if you removed a human being's natural incentive to work—his initiative, his spontaneous creative energy— and replaced it with external motivation and coercion, he would become a lazy and careless worker." (Le Guin 1974: p. 82). This perspective echoes the views of Odo, who conceived of many of the key ideas and practices reflected on Anarres. The belief in freedom to use one's time how one chooses to (including to shun social interaction all together, as some do in this world) is ubiquitous on Anarres. This ideal of freedom is part of a broader web of widely held cultural Odonian beliefs and values, and is juxtaposed with a complex system for allocating people's time to socially beneficial tasks. The largely computerized system for job assignment produces a state of affairs which, while ostensibly free, in fact contributes to a different form of oppression—not by "profiteers" (a term the Anarresi use pejoratively) but by the judgment of one's peers. When Oiee asks Shevek, "No one ever defies it [work assignment]?," Shevek replies, "Perhaps not often enough." (150)

Ahmed's analysis of willfulness helps us to understand this situation. On Anarres, while freedom is a central part of the society's ideology, in practice, dynamics of willingness or willfulness come into play depending on the particular subject. If someone's will conforms to the general will, then they have a smoother time of things, as Ahmed notes. Yet social pressure, as on Anarres, can militate against the use of will that goes against the general will. Even if the general will is ostensibly or apparently utopian with respect to a certain conception of justice, this does not necessarily prevent it from, over time, crystallizing into an oppressive ideology through the sorts of simplifications and lack of self-criticism that Kitch warns about.

In *The Dispossessed*, the ostensible lack of laws on Anarres is misleading—the belief among Odonians that their utopia has been realized blinds them to the limits of their world. One of these limits is its intellectual and ideological stasis—demonstrated, for example, by the fierce initial opposition to Shevek's scientific and political ideas. Shevek, recognizing some of these limits, exhibits willfulness in Ahmed's sense by actively cultivating controversy through forming the Syndicate of Initiative and being deliberately provocative in his interactions with others.

Like the feminist killjoys Ahmed discusses, Shevek recognizes a problem in his society and resists it with his words and deeds, even when doing so may be uncomfortable to himself and others. Le Guin's work is thus a key contribution to the willfulness archive in two respects. First, it gives numerous examples of practices and ideas that could be used to resist injustice, as well as some of their possible limitations. Second, it reminds us that the process of realizing justice *to its fullest possible extent* never ends, and that we should not put stock in any particular vision of what utopia consists of (for example, control over one's time). In short, the need for willfulness is not eroded even in a society that claims to value it (or freedom) above all else, which suggests that a properly defined utopianism needs to account explicitly for this processual nature of justice's realization and the risks of negative side-effects in realizing particular aspects of justice, calling for new forms of subjectivity and utopian visions in response to the new ideologies (which could once have been utopias, pace Mannheim). Thus, learning from Ahmed, *The Dispossessed*, and the utopian literature more broadly, we see reasons not to focus exclusively on the outcomes (e.g. the elimination of paid work) but also the means of self-criticism afforded by particular institutions, practices, and ideas, and the processes available for correcting mistakes of utopian projects of the past. With this perspective in mind, we turn to the final section of this essay, which will analyze visions of the future of AI and work from the perspective of utopianism.

AI and the Future of Justice

As previously noted, goals and visions such as the "eradication of poverty" are mainstream in contemporary discussions of AI, such as in the aforementioned open letter (Future of Life Institute 2015). These visions posit a central role for technology in the realization of justice, and thus the attainment of utopia. Among other issues with these visions, they often eschew serious consideration of the dense interconnections of humans and technologies in the course of social change. Rather, proponents of AI as a path to utopia often instead conceive of technologies as proceeding mostly autonomously in a way that can be analyzed in isolation and then straightforwardly technocratically governed to realize social welfare, without addressing deep social inequities and other social issues. A paradigmatic example of this perspective is philosopher Nick Bostrom's (2014) book *Superintelligence*, which suggests (not without significant underlying analysis, but still troublingly) that a highly secretive agency formed by the world's governments, using extensive internal and external surveillance, might be the best means of governing advanced AI technologies, which in turn, he argues, would provide the foundation for eliminating disease, poverty, environmental degradation, etc. which are conceived as reducible to insufficient intelligence

applied to those problems. With regard to questions of justice, Bostrom argues that AI-based technologies should be designed and managed in the common interests of humanity, and that broadly shared values should be reflected in the design. Beyond these stipulations, however, no serious attention is given in this work to the fact of deep diversity of values around the world, the problems of entrusting states with humanity's future, or ensuring that the management of such technologies is amenable to the sort of deep and broad-based criticism Kitch calls for and Ahmed's work suggests is needed.

Beyond Bostrom's book, the content of common (purportedly) utopian visions for the future of AI in relation to work in particular demands sustained critique and constructive improvement if justice in is in fact to be realized in the future in the face of these socio-technological affordances. Without such critique, Kitch's concerns about utopianism may be vindicated in the utopias of the future, and the sorts of problems faced by Shevek in *The Dispossessed* may be reproduced in a world claiming to be utopian but falling far short. Two weaknesses of contemporary utopian AI visions stand out in light of the foregoing analysis, which I'll cover in turn after briefly summarizing the content of these visions. Bostrom's work is representative of a large stream of research on superintelligence safety issues—both already existing work and that which may soon be funded by the Future of Life Institute, which echoes his concerns and was funded by Elon Musk, who references Bostrom's book regularly. However, there is as yet only an emerging tapestry of work on the future of work viz-a-viz AI, and no consensus exists on how to grapple with the socio-technological opportunities and risks afforded by AI and robotics in the coming decades. One recent work in this domain, *The Second Machine Age* by Andrew McAfee and Erik Brynjolfsson (2014), argues against the provision of a basic income as an approach for widely sharing the productivity benefits afforded by advanced technologies, conceiving of work as something to be actively encouraged through e.g. better education and immigration policies, more entrepreneurship, and so forth. Their mantra is "racing with the machines," i.e. coupling the "best" of human and machine capabilities in redesigning jobs and processes to be more productive and to enable continued human participation in the economy for the foreseeable future. Others such as Martin Ford in his book *Rise of the Robots* (2015) advocates for a basic income and sees, echoing John Maynard Keynes, the creation of a "leisure society" as a viable path towards utopia, reflecting his greater optimism about people's ability to find new ways of using their time in a "post-work" world. Whether in advocating for a "post-work" like Ford or a "post-bad work" vision of the future like McAfee and Brynjolfsson, these utopias and many others put forward along similar lines share two common weaknesses.

First, the question of utopian subjectivities is given short thrift in many imaginings of the future of AI and society. Bostrom's attention to this topic in *Superintelligence,* for example, consists merely of exhortations for humans to be more

rational and cooperative in order to realize the technologically-derived benefits he envisions. However, as Ahmed's *Willful Subjects* argues and *The Dispossessed* demonstrates, there is an integral connection between processes of subject formation on the one hand and utopian processes and outcomes on the other. Indeed, the language used by Anaressi, in which they refer to play and work with the same word, is representative of the sort of deeper reflection on our future possible relationships with each other and with technologies that need to be envisioned. McAfee and Brynjolfsson's optimistic orientation toward technology is paired with a pessimistic orientation toward human subjects—namely, that a never ending supply of jobs is the solution to our problems---a more imaginative, forward-thinking, or humane relationship between people and their work and communities is beyond the scope of their vision. In Mannheim's terms, the capitalist ideology that infuses their work needs to be opened up to a more robust utopian critique that takes subjectivity seriously. Even explicit advocates of the utopian potential of a basic income, such as Phillipe van Parijs, give short thrift to such considerations. In his paper, "The Universal Income: Why Utopian Thinking Matters, and How Sociologists Can Contribute to It," he writes, "Don't worry: I'm not going to waste any of your time describing the idyllic life of the inhabitants of basic income utopia." (Van Parijs 2013, p. 174). It is notable that not only does Van Parijs eschew the subjective dimensions of utopia, but he explicitly characterizes them as a waste of time.

Second, "work" in visions of a "post-work" future typically conflate work with paid work, and more specifically particular idealized kinds of work such as manufacturing, management and development of information technology, etc. that are often dominated by men. In contrast, care work, reproductive work, and other forms of work more associated with women rarely figure in these visions in any significant way. Little if any consideration is given to e.g. the question of what role AI and robotics, if any, would play in reshaping these domains of life, if this happening would be desirable, and how to account for the gendered and raced nature of so much work in the world today in formulating approaches to the governance of AI and robotics. In *Border as Method, Or, The Multiplication of Labor*, Mezzadra and Neilson develop a nuanced account such issues in relation to the proliferation and heterogenization of borders in recent years. They write:

> "Labor was multiplied through these processes in at least three important ways. It was first *intensified*, in the sense that its tendency to colonize the entire life of laboring subjects became even more pronounced than before. Second, it was internally *diversified*, according to a process already identified by Marx in his analysis of the creation of relative surplus value in the *Grundrisse*, which continuously pushes capital beyond the division of labor and toward "the development of a constantly expanding and more comprehensive system of different kinds of labor, different kinds of production, to which a constantly expanding and constantly enriched system of needs

corresponds (Marx 1973, 409). Third, it was *heterogenized*, as far as legal and social regimes of its organization are

concerned." (Mezzadra and Neilson 2013, p. 88).

Such considerations are entirely absent from most writings on AI and the future of work, suggesting the critical need for

evaluating not only the content of these utopian visions and the assumptions underlying them, but the extent to which they

can even be properly conceived as utopias. Again drawing on Mannheim's work, insofar as "racing with the machines" can

be construed as a possible form of future intensification, diversification, and heterogenization of work, then it's not obvious

that "post work" futures are really distinct from contemporary capitalist ideology. This question demands critical

engagement if the potential Bloch, Gordon, and others see in utopianism is to be realized in practice, and if Kitch's concerns

are to be substantively addressed.

Conclusion

Through an analysis of the works of Gordon, Bloch, Mannheim, we saw the complex and evolving nature of

utopianism and the reasons for its political contestation. In Kitch's work, we found serious critiques of utopianism based on

an extrapolation from particular cases, which raised fundamental questions and risks regarding utopianism but did not

succeed in undermining it in its entirety. In Ahmed's *Willful Subjects* and Le Guin's *The Dispossessed*, we saw reasons to

conceive of utopia and justice not merely in terms of material and other outcomes but also in terms of processes and

subjectivities. Finally, in (questionably) utopian visions of the future of work in relation to AI, we found the urgent

relevance of questions raised by Kitch, Ahmed, and others in the context of contemporary utopian visions.

The outlook for utopia and the future of justice suggested by this analysis, then, is as follows. Utopianism is a vast

and flexible body of theory and practice oriented towards the realization of justice, which eludes a crisp definition or

devastating critique. Through engagement with history, contemporary political struggles, and science fiction, the cultivation

and use of a willfulness archive as suggested by Ahmed, the self-criticism and other dimensions of realism suggested by

Kitch, and attending deeply to subjectivities and processes of justice, we can attempt to improve the caliber of utopian

thought and practice and utopianism more generally. Finally, Bloch's analysis reminds of the centrality of human choice and

possibility in moving toward the future. While this vastness of possibility has its risks, our human capabilities, longings, and

potential for growth (both individually and collectively) can help us to think further ahead, think more broadly today, and

think deeper into the past and the space of possibilities represented by science fiction. While none of this guarantees

justice's realization, it provides reason for hope that the utopianism of the future can be better than the utopianism of the

present and the past, as we learn more about the difficulties and possibilities of justice.

Works Cited

Ahmed, S. 2014. *Willful Subjects*. Duke University Press: Durham.

Bloch, E. 1954. *The Principle of Hope*, *Volume 1*. MIT Press: Cambridge.

Bostrom, 2014. *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press: Oxford.

Brynjolfsson, E. & McAfee, A. 2014. *The Second Machine Age*. MIT Press: Cambridge.

Ford, M. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books: New York.

Future of Life Institute. 2015. "Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter," Accessable online at: http://futureoflife.org/misc/open_letter

Gordon, A. 2004. *Keeping Good Time: Reflections on Knowledge, Power, and People*. Paradigm Publishers: Boulder.

Kitch, S. 2000. *Higher Ground: From Utopianism to Realism in American Feminist Thought and Theory*. University of Chicago Press: Chicago.

Le Guin, U. 1974. *The Dispossessed: An Ambiguous Utopia*. Harper and Row: London.

Mannheim, K. 1932. *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. Harvest Book: New York.

Mezzadra, S. & Neilson, B. 2013. *Border as Method, or, the Multiplication of Labor.* Duke University Press: Durham.

Robinson, C. 2001. *An Anthropology of Marxism*. Ashgate: Sydney.

Van Parijs, P. 2013. "The Universal Basic Income: Why Utopian Thinking Matters, and How Sociologists Can Contribute to It." *Politics & Society*, 41(2).