Transhumanist Values

NICK BOSTROM

Oxford University, Faculty of Philosophy
Homepage: www.nickbostrom.com

[in *Ethical Issues for the 21st Century*, ed. Frederick Adams (Philosophical Documentation Center Press, 2003); reprinted in *Review of Contemporary Philosophy*, Vol. 4, May (2005)]
[pdf]

1. What is Transhumanism?

Transhumanism is a loosely defined movement that has developed gradually over the past two decades.[1] It promotes an interdisciplinary approach to understanding and evaluating the opportunities for enhancing the human condition and the human organism opened up by the advancement of technology. Attention is given to both present technologies, like genetic engineering and information technology, and anticipated future ones, such as molecular nanotechnology and artificial intelligence.

The enhancement options being discussed include radical extension of human health-span, eradication of disease, elimination of unnecessary suffering, and augmentation of human intellectual, physical, and emotional capacities. Other transhumanist themes include space colonization and the possibility of creating superintelligent machines, along with other potential developments that could profoundly alter the human condition. The ambit is not limited to gadgets and medicine, but encompasses also economic, social, institutional designs, cultural development, and psychological skills and techniques.

Transhumanists view human nature as a work-in-progress, a half-baked beginning that we can learn to remold in desirable ways. Current humanity need not be the endpoint of evolution. Transhumanists hope that by responsible use of science, technology, and other rational means we shall eventually manage to become posthuman, beings with vastly greater capacities than present human beings have.

Some transhumanists take active steps to increase the probability that they personally will survive long enough to become posthuman, for example by choosing a healthy lifestyle or by making provisions for having themselves cryonically suspended in case of de-animation.[2] In contrast to many other ethical outlooks, which in practice often reflect a reactionary attitude to new technologies, the transhumanist view is guided by an evolving vision to take a more proactive approach to technology policy. This vision, in broad strokes, is to create the opportunity to live much longer and healthier lives, to enhance our memory and other intellectual faculties, to refine our emotional experiences and increase our subjective sense of well-being, and generally to achieve a greater degree

of control over our own lives. This affirmation of human potential is offered as an alternative to customary injunctions against playing God, messing with nature, tampering with our human essence, or displaying punishable hubris.

Transhumanism does not entail technological optimism. While future technological capabilities carry immense potential for beneficial deployments, they also could be misused to cause enormous harm, ranging all the way to the extreme possibility of intelligent life becoming extinct. Other potential negative outcomes include widening social inequalities or a gradual erosion of the hard-to-quantify assets that we care deeply about but tend to neglect in our daily struggle for material gain, such as meaningful human relationships and ecological diversity. Such risks must be taken very seriously, as thoughtful transhumanists fully acknowledge.[3]

Transhumanism has roots in secular humanist thinking, yet is more radical in that it promotes not only traditional means of improving human nature, such as education and cultural refinement, but also direct application of medicine and technology to overcome some of our basic biological limits.

2. Human limitations

The range of thoughts, feelings, experiences, and activities accessible to human organisms presumably constitute only a tiny part of what is possible. There is no reason to think that the human mode of being is any more free of limitations imposed by our biological nature than are those of other animals. In much the same way as Chimpanzees lack the cognitive wherewithal to understand what it is like to be human – the ambitions we humans have, our philosophies, the complexities of human society, or the subtleties of our relationships with one another, so we humans may lack the capacity to form a realistic intuitive understanding of what it would be like to be a radically enhanced human (a “posthuman”) and of the thoughts, concerns, aspirations, and social relations that such humans may have.

Our own current mode of being, therefore, spans but a minute subspace of what is possible or permitted by the physical constraints of the universe (see Figure 1). It is not farfetched to suppose that there are parts of this larger space that represent extremely valuable ways of living, relating, feeling, and thinking.

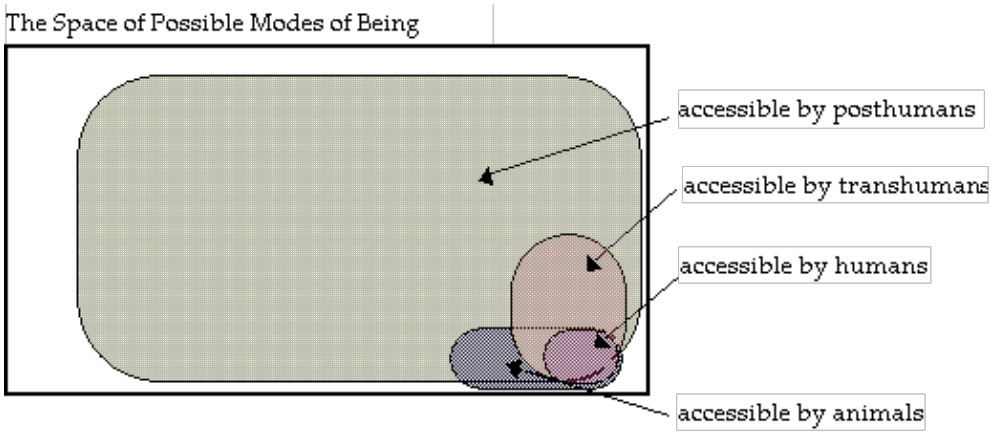

Figure 1. We aint seen nothin' yet (not drawn to scale). The term "transhuman" denotes transitional beings, or moderately enhanced humans, whose capacities would be somewhere between those of unaugmented humans and full-blown posthumans. (A transhumanist, by contrast, is simply somebody who accepts transhumanism.)

The limitations of the human mode of being are so pervasive and familiar that we often fail to notice them, and to question them requires manifesting an almost childlike naiveté. Let consider some of the more basic ones.

Lifespan. Because of the precarious conditions in which our Pleistocene ancestors lived, the human lifespan has evolved to be a paltry seven or eight decades. This is, from many perspectives, a rather short period of time. Even tortoises do better than that.

We don't have to use geological or cosmological comparisons to highlight the meagerness of our allotted time budgets. To get a sense that we might be missing out on something important by our tendency to die early, we only have to bring to mind some of the worthwhile things that we could have done or attempted to do if we had had more time. For gardeners, educators, scholars, artists, city planners, and those who simply relish observing and participating in the cultural or political variety shows of life, three scores and ten is often insufficient for seeing even one major project through to completion, let alone for undertaking many such projects in sequence.

Human character development is also cut short by aging and death. Imagine what might have become of a Beethoven or a Goethe if they had still been with us today. Maybe they would have developed into rigid old grumps interested exclusively in conversing about the achievements of their youth. But maybe, if they had continued to enjoy health and youthful vitality, they would have continued to grow as men and artists, to reach levels of maturity that we can barely imagine. We certainly cannot rule that out based on what we know today. Therefore, there is at least a serious possibility of there being something very precious outside the human sphere. This constitutes a reason to pursue the means that will let us go there and find out.

Intellectual capacity. We have all had moments when we wished we were a little smarter. The three-pound, cheese-like thinking machine that we lug around in our skulls can do some neat tricks, but it also has

significant shortcomings. Some of these – such as forgetting to buy milk or failing to attain native fluency in languages you learn as an adult – are obvious and require no elaboration. These shortcomings are inconveniences but hardly fundamental barriers to human development.

Yet there is a more profound sense in the constraints of our intellectual apparatus limit our modes of our mentation. I mentioned the Chimpanzee analogy earlier: just as is the case for the great apes, our own cognitive makeup may foreclose whole strata of understanding and mental activity. The point here is not about any logical or metaphysical impossibility: we need not suppose that posthumans would not be Turing computable or that they would have concepts that could not be expressed by any finite sentences in our language, or anything of that sort. The impossibility that I am referring to is more like the impossibility for us current humans to visualize an 200-dimensional hypersphere or to read, with perfect recollection and understanding, every book in the Library of Congress. These things are impossible for us because, simply put, we lack the brainpower. In the same way, may lack the ability to intuitively understand what being a posthuman would be like or to grok the playing field of posthuman concerns.

Further, our human brains may cap our ability to discover philosophical and scientific truths. It is possible that failure of philosophical research to arrive at solid, generally accepted answers to many of the traditional big philosophical questions could be due to the fact that we are not smart enough to be successful in this kind of enquiry. Our cognitive limitations may be confining us in a Platonic cave, where the best we can do is theorize about "shadows", that is, representations that are sufficiently oversimplified and dumbed-down to fit inside a human brain.

Bodily functionality. We enhance our natural immune systems by getting vaccinations, and we can imagine further enhancements to our bodies that would protect us from disease or help us shape our bodies according to our desires (e.g. by letting us control our bodies' metabolic rate). Such enhancements could improve the quality of our lives.

A more radical kind of upgrade might be possible if we suppose a computational view of the mind. It may then be possible to upload a human mind to a computer, by replicating *in silico* the detailed computational processes that would normally take place in a particular human brain.[4] Being an upload would have many potential advantages, such as the ability to make back-up copies of oneself (favorably impacting on one's life-expectancy) and the ability to transmit oneself as information at the speed of light. Uploads might live either in virtual reality or directly in physical reality by controlling a robot proxy.

Sensory modalities, special faculties and sensibilities. The current human sensory modalities are not the only possible ones, and they are certainly not as highly developed as they could be. Some animals have sonar, magnetic orientation, or sensors for electricity and vibration; many have a much keener sense of smell, sharper eyesight, etc. The range of possible sensory modalities is not limited to those we find in the animal

kingdom. There is no fundamental block to adding say a capacity to see infrared radiation or to perceive radio signals and perhaps to add some kind of telepathic sense by augmenting our brains with suitably interfaced radio transmitters.

Humans also enjoy a variety of special faculties, such as appreciation of music and a sense of humor, and sensibilities such as the capacity for sexual arousal in response to erotic stimuli. Again, there is no reason to think that what we have exhausts the range of the possible, and we can certainly imagine higher levels of sensitivity and responsiveness.

Mood, energy, and self-control. Despite our best efforts, we often fail to feel as happy as we would like. Our chronic levels of subjective well-being seem to be largely genetically determined. Life-events have little long-term impact; the crests and troughs of fortune push us up and bring us down, but there is little long-term effect on self-reported well-being. Lasting joy remains elusive except for those of us who are lucky enough to have been born with a temperament that plays in a major key.

In addition to being at the mercy of a genetically determined setpoint for our levels of well-being, we are limited in regard to energy, will-power, and ability to shape our own character in accordance with our ideals. Even such "simple" goals as losing weight or quitting smoking prove unattainable to many.

Some subset of these kinds of problems might be necessary rather than contingent upon our current nature. For example, we cannot both have the ability easily to break any habit and the ability to form stable, hard-to-break habits. (In this regard, the best one can hope for may be the ability to easily get rid of habits we didn't deliberately choose for ourselves in the first place, and perhaps a more versatile habit-formation system that would let us choose with more precision when to acquire a habit and how much effort it should cost to break it.)

3. The core transhumanist value: exploring the posthuman realm

The conjecture that there are greater values than we can currently fathom does not imply that values are not defined in terms of our current dispositions. Take, for example, a dispositional theory of value such as the one described by David Lewis.[5] According to Lewis's theory, something is a value for you if and only if you would want to want it if you were perfectly acquainted with it and you were thinking and deliberating as clearly as possible about it. On this view, there may be values that we do not currently want, and that we do not even currently want to want, because we may not be perfectly acquainted with them or because we are not ideal deliberators. Some values pertaining to certain forms of posthuman existence may well be of this sort; they may be values for us now, and they may be so in virtue of our current dispositions, and yet we may not be able to fully appreciate them with our current limited deliberative capacities and our lack of the receptive

faculties required for full acquaintance with them. This point is important because it shows that the transhumanist view that we ought to explore the realm of posthuman values does not entail that we should forego our current values. The posthuman values can be our current values, albeit ones that we have not yet clearly comprehended. Transhumanism does not require us to say that we should favor posthuman beings over human beings, but that the right way of favoring human beings is by enabling us to realize our ideals better and that some of our ideals may well be located outside the space of modes of being that are accessible to us with our current biological constitution.

We can overcome many of our biological limitations. It is possible that there are some limitations that are impossible for us to transcend, not only because of technological difficulties but on metaphysical grounds. Depending on what our views are about what constitutes personal identity, it could be that certain modes of being, while possible, are not possible for us, because any being of such a kind would be so different from us that they could not be us. Concerns of this kind are familiar from theological discussions of the afterlife. In Christian theology, some souls will be allowed by God to go to heaven after their time as corporal creatures is over. Before being admitted to heaven, the souls would undergo a purification process in which they would lose many of their previous bodily attributes. Skeptics may doubt that the resulting minds would be sufficiently similar to our current minds for it to be possible for them to be the same person. A similar predicament arises within transhumanism: if the mode of being of a posthuman being is radically different from that of a human being, then we may doubt whether a posthuman being could be the same person as a human being, even if the posthuman being originated from a human being.

We can, however, envision many enhancements that would not make it impossible for the post-transformation someone to be the same person as the pre-transformation person. A person could obtain quite a bit of increased life expectancy, intelligence, health, memory, and emotional sensitivity, without ceasing to exist in the process. A person's intellectual life can be transformed radically by getting an education. A person's life expectancy can be extended substantially by being unexpectedly cured from a lethal disease. Yet these developments are not viewed as spelling the end of the original person. In particular, it seems that modifications that add to a person's capacities can be more substantial than modifications that subtract, such as brain damage. If most of someone currently is, including her most important memories, activities, and feelings, is preserved, then adding extra capacities on top of that would not easily cause the person to cease to exist.

Preservation of personal identity, especially if this notion is given a narrow construal, is not everything. We can value other things than ourselves, or we might regard it as satisfactory if some parts or aspects of ourselves survive and flourish, even if that entails giving up some parts of ourselves such that we no longer count as being the same person. Which parts of ourselves we might be willing to sacrifice may not

become clear until we are more fully acquainted with the full meaning of the options. A careful, incremental exploration of the posthuman realm may be indispensable for acquiring such an understanding, although we may also be able to learn from each other's experiences and from works of the imagination.

Additionally, we may favor future people being posthuman rather than human, if the posthumans would lead lives more worthwhile than the alternative humans would. Any reasons stemming from such considerations would not depend on the assumption that we ourselves could become posthuman beings.

Transhumanism promotes the quest to develop further so that we can explore hitherto inaccessible realms of value. Technological enhancement of human organisms is a means that we ought to pursue to this end. There are limits to how much can be achieved by low-tech means such as education, philosophical contemplation, moral self-scrutiny and other such methods proposed by classical philosophers with perfectionist leanings, including Plato, Aristotle, and Nietzsche, or by means of creating a fairer and better society, as envisioned by social reformists such as Marx or Martin Luther King. This is not to denigrate what we can do with the tools we have today. Yet ultimately, transhumanists hope to go further.

4. Basic conditions for realizing the transhumanist project

If this is the grand vision, what are the more particular objectives that it translates into when considered as a guide to policy?

What is needed for the realization of the transhumanist dream is that technological means necessary for venturing into the posthuman space are made available to those who wish to use them, and that society be organized in such a manner that such explorations can be undertaken without causing unacceptable damage to the social fabric and without imposing unacceptable existential risks.

Global security. While disasters and setbacks are inevitable in the implementation of the transhumanist project (just as they are if the transhumanist project is not pursued), there is one kind of catastrophe that must be avoided at any cost:

> *Existential risk* – one where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential.[6]

Several recent discussions have argued that the combined probability of the existential risks is very substantial.[7] The relevance of the condition of existential safety to the transhumanist vision is obvious: if we go extinct or permanently destroy our potential to develop further, then the transhumanist core value will not be realized. Global security is the most

fundamental and nonnegotiable requirement of the transhumanist project.

Technological progress. That technological progress is generally desirable from a transhumanist point of view is also self-evident. Many of our biological shortcomings (aging, disease, feeble memories and intellects, a limited emotional repertoire and inadequate capacity for sustained well-being) are difficult to overcome, and to do so will require advanced tools. Developing these tools is a gargantuan challenge for the collective problem-solving capacities of our species. Since technological progress is closely linked to economic development, economic growth – or more precisely, productivity growth – can in some cases serve as a proxy for technological progress. (Productivity growth is, of course, only an imperfect measure of the relevant form of technological progress, which, in turn, is an imperfect measure of overall improvement, since it omits such factors as equity of distribution, ecological diversity, and quality of human relationships.)

The history of economic and technological development, and the concomitant growth of civilization, is appropriately regarded with awe, as humanity's most glorious achievement. Thanks to the gradual accumulation of improvements over the past several thousand years, large portions of humanity have been freed from illiteracy, life-expectancies of twenty years, alarming infant-mortality rates, horrible diseases endured without palliatives, and periodic starvation and water shortages. Technology, in this context, is not just gadgets but includes all instrumentally useful objects and systems that have been deliberately created. This broad definition encompasses practices and institutions, such as double-entry accounting, scientific peer-review, legal systems, and the applied sciences.

Wide access. It is not enough that the posthuman realm be explored by someone. The full realization of the core transhumanist value requires that, ideally, everybody should have the opportunity to become posthuman. It would be sub-optimal if the opportunity to become posthuman were restricted to a tiny elite.

There are many reasons for supporting wide access: to reduce inequality; because it would be a fairer arrangement; to express solidarity and respect for fellow humans; to help gain support for the transhumanist project; to increase the chances that you will get the opportunity to become posthuman; to increase the chances that those you care about can become posthuman; because it might increase the range of the posthuman realm that gets explored; and to alleviate human suffering on as wide a scale as possible.

The wide access requirement underlies the *moral urgency* of the transhumanist vision. Wide access does not argue for holding back. On the contrary, other things being equal, it is an argument for moving forward as quickly as possible. 150,000 human beings on our planet die every day, without having had any access to the anticipated enhancement technologies that will make it possible to become posthuman. The sooner this technology develops, the fewer people will have died without access.

Consider a hypothetical case in which there is a choice between (a) allowing the current human population to continue to exist, and (b) having it instantaneously and painlessly killed and replaced by six billion new human beings who are very similar but non-identical to the people that exist today. Such a replacement ought to be strongly resisted on moral grounds, for it would entail the involuntary death of six billion people. The fact that they would be replaced by six billion newly created similar people does not make the substitution acceptable. Human beings are not disposable. For analogous reasons, it is important that the opportunity be become posthuman is made available to as many humans as possible, rather than having the existing population merely supplemented (or worse, replaced) by a new set of posthuman people. The transhumanist ideal will be maximally realized only if the benefits of technologies are widely shared and if they are made available as soon as possible, preferably within our lifetime.

5. Derivative values

From these specific requirements flow a number of derivative transhumanist values that translate the transhumanist vision into practice. (Some of these values may also have independent justifications, and transhumanism does not imply that that the list of values provided below is exhaustive.)

To start with, transhumanists typically place emphasis on individual freedom and individual choice in the area of enhancement technologies. Humans differ widely in their conceptions of what their own perfection or improvement would consist in. Some want to develop in one direction, others in different directions, and some prefer to stay the way they are. It would neither be morally unacceptable for anybody to impose a single standard to which we would all have to conform. People should have the right to choose which enhancement technologies, if any, they want to use. In cases where individual choices impact substantially on other people, this general principle may need to be restricted, but the mere fact that somebody may be disgusted or morally affronted by somebody else's using technology to modify herself would not normally a legitimate ground for coercive interference. Furthermore, the poor track record of centrally planned efforts to create better people (e.g. the eugenics movement and Soviet totalitarianism) shows that we need to be wary of collective decision-making in the field of human modification.

Another transhumanist priority is to put ourselves in a better position to make wise choices about where we are going. We will need all the wisdom we can get when negotiating the posthuman transition. Transhumanists place a high value on improvements in our individual and collective powers of understanding and in our ability to implement responsible decisions. Collectively, we might get smarter and more informed through such means as scientific research, public debate and open discussion of the future, information markets[8], collaborative information filtering[9]. On an individual level, we can benefit from education, critical thinking, open-mindedness, study techniques,

information technology, and perhaps memory- or attention-enhancing drugs and other cognitive enhancement technologies. Our ability to implement responsible decisions can be improved by expanding the rule of law and democracy on the international plane. Additionally, artificial intelligence, especially if and when it reaches human-equivalence or greater, could give an enormous boost to the quest for knowledge and wisdom.

Given the limitations of our current wisdom, a certain epistemic tentativeness is appropriate, along with a readiness to continually reassess our assumptions as more information becomes available. We cannot take for granted that our old habits and beliefs will prove adequate in navigating our new circumstances.

Global security can be improved by promoting international peace and cooperation, and by strongly counteracting the proliferation of weapons of mass destruction. Improvements in surveillance technology may make it easier to detect illicit weapons programs. Other security measures might also be appropriate to counteract various existential risks. More studies on such risks would help us get a better understanding of the long-term threats to human flourishing and of what can be done to reduce them.

Since technological development is necessary to realize the transhumanist vision, entrepreneurship, science, and the engineering spirit are to be promoted. More generally, transhumanists favor a pragmatic attitude and a constructive, problem-solving approach to challenges, preferring methods that experience tells us give good results. They think it better to take the initiative to "do something about it" rather than sit around complaining. This is one sense in which transhumanism is optimistic. (It is not optimistic in the sense of advocating an inflated belief in the probability of success or in the Panglossian sense of inventing excuses for the shortcomings of the status quo.)

Transhumanism advocates the well-being of all sentience, whether in artificial intellects, humans, and non-human animals (including extraterrestrial species, if there are any). Racism, sexism, speciesism, belligerent nationalism and religious intolerance are unacceptable. In addition to the usual grounds for deeming such practices objectionable, there is also a specifically transhumanist motivation for this. In order to prepare for a time when the human species may start branching out in various directions, we need to start now to strongly encourage the development of moral sentiments that are broad enough encompass within the sphere of moral concern sentiences that are constituted differently from ourselves.

Finally, transhumanism stresses the moral urgency of saving lives, or, more precisely, of preventing involuntary deaths among people whose lives are worth living. In the developed world, aging is currently the number one killer. Aging is also biggest cause of illness, disability and dementia. (Even if all heart disease and cancer could be cured, life expectancy would increase by merely six to seven years.) Anti-aging

medicine is therefore a key transhumanist priority. The goal, of course, is to radically extent people's active health-spans, not to add a few extra years on a ventilator at the end of life.

Since we are still far from being able to halt or reverse aging, cryonic suspension of the dead should be made available as an option for those who desire it. It is possible that future technologies will make it possible to reanimate people who have cryonically suspended.[10] While cryonics might be a long shot, it definitely carries better odds than cremation or burial.

The table below summarizes the transhumanist values that we have discussed.

TABLE OF TRANSHMANIST VALUES

Core Value

- Having the opportunity to explore the transhuman and posthuman realms

Basic Conditions

- Global security
- Technological progress
- Wide access

Derivative Values

- Nothing wrong about "tampering with nature"; the idea of *hubris* rejected
- Individual choice in use of enhancement technologies; morphological freedom
- Peace, international cooperation, anti-proliferation of WMDs
- Improving understanding (encouraging research and public debate; critical thinking; open-mindedness, scientific inquiry; open discussion of the future)
- Getting smarter (individually; collectively; and develop machine intelligence)
- Philosophical fallibilism; willingness to reexamine assumptions as we go along
- Pragmatism; engineering- and entrepreneur-spirit; science
- Diversity (species, races, religious creeds, sexual orientations, life styles, etc.)
- Caring about the well-being of all sentience
- Saving lives (life-extension, anti-aging research, and cryonic

Figure 2. Table of transhumanist values

References

Bostrom, N. (2002). "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards." *Journal of Evolution and Technology*, 9.

Bostrom, N. (2003). "Human Genetic Enhancements: A Transhumanist Perspective." *Journal of Value Inquiry*, Forthcoming.

Bostrom, N., et al. (1999). *The Transhumanist FAQ*.

http://www.transhumanism.org/resources/FAQv21.pdf

Chislenko, A. (1997). *Automated Collaborative Filtering and Semantic Transports*. http://www.lucifer.com/%7Esasha/articles/ACF.html

Drexler, K. E. *Engines of Creation: The Coming Era of Nanotechnology*. (Anchor Books: New York, 1986).

Ettinger, R. *The Prospect of Immortality*. (Doubleday: New York, 1964).

Hanson, R. (1995). "Could Gambling Save Science? Encouraging an Honest Consensus." *Social Epistemology*, 9:1: 3-33.

Hughes, J. (2001). "The Future of Death: Cryonics and the Telos of Liberal Individualism." *Journal of Evolution and Technology*, 6.

Leslie, J. *The End of the World: The Science and Ethics of Human Extinction*. (Routledge: London, 1996).

Lewis, D. (1989). "Dispositional Theories of Value." *Proceedings of the Aristotelian Society Supp.*, 63: 113-37.

Merkle, R. (1994). "The Molecular Repair of the Brain." *Cryonics*, 15(1 and 2).

Moravec, H. *Mind Children*. (Harvard University Press: Harvard, 1989).

Rees, M. *Our Final Hour: A Scientist's Warning*. (Basic Books: New York, 2003).

---

[1] (Bostrom et al. 1999; Bostrom 2003)
[2] (Ettinger 1964; Hughes 2001)
[3] (Bostrom 2002)
[4] (Drexler 1986; Moravec 1989)
[5] (Lewis 1989)
[6] (Bostrom 2002)
[7] (Leslie 1996; Bostrom 2002; Rees 2003)
[8] (Hanson 1995)
[9] (Chislenko 1997)
[10] (Ettinger 1964; Drexler 1986; Merkle 1994)