The revolution is back

(Turing, functional realization and computational description)

Reza Negarestani

I philosophically endorse computationalism and even more so I am an ardent proponent of functionalism. I think—and I am fully prepared to defend this controversial claim—that a philosopher cannot intellectually survive without endorsing functionalism, at least one of its many varieties (strongly normative [Hegel, Brandom], normative-materially constrained [Sellars] or strongly mechanistic [Bechtel]). To this extent, what I would like to briefly address is the significance of the functionalist account of the human mind, or more broadly speaking, the functionalist account of the rational agency. In this respect, I take side with Alan Turing's response to *Arguments from Various Disabilities* (AVD) where he challenges the common forms of rejecting the possibility of the functional realization of the human mind in different substrates—for instance, in machines.

Machines cannot think, machines cannot have emotions, machines cannot be purposeful, they cannot be proactive and so forth: Turing enumerates these under what he calls arguments from various disabilities, it is sort of *straw machine argument* that is baseless and precarious. It is more a fruit of our psychological fears and residual theological approaches to the universe and ourselves than the result of sound arguments.

The mind-preservationist is a person who believes that the mind cannot be functionally realized and implemented in different substrates. He is a person who not only rejects the functionalist realization of the mind but also as a result yields to a form of vitalism or ineffability of the human mind. The mind-preservationist always attempts to see the machine's capacities from the perspective of an endemic disability. But if what the mind-preservationist really dismisses is not the machine as such but is the functional realization of the mind implemented in the machine, then what he actually denies is not the machine per se but the mind itself. Or more accurately, what the mind-preservationist ends up rejecting is the possibility of mapping the mind's functions, the possibility of modeling it, defining and objectifying it. In this sense, machine-denialism is simply an excuse for denying what the mind is and what it can be. Correspondingly, disavowing the pursuit of understanding the mind coincides with acting against the evolution of the mind, since from a pragmatic-functional viewpoint the understanding of the meaning of the mind is inseparable from how the mind can be defined, reconstructed and modified in different contexts. Therefore, if we lack the definition of the mind which is itself a map for its realization and objectification, then how can we so readily rule out the possibility of a machine furnished with a mind? The mind-preservationist, accordingly, has a double standard when it comes to recognizing the mind as both the measure and the object of his critique. He says the machine cannot engage in mental activities as if he possesses the map of the mind. However, if he does not know what constitutes activities of the mind, which is to say, if he does not possess the functional map of the mind, then he cannot approach the functional account of the mind (that is, a mind realized by a

different set of realizers and implemented in an environment different from its natural-biological habitat) from the perspective of an intrinsic disability.

If you don't know what the mind is then how can you claim the machine cannot possibly have a mind? With the understanding that the 'what' posed in this question is the very map of the mind's functional realizability that can be implemented in machines. Here 'what' can be described functionally as those *activities* which define what the mind is. The mind is therefore described as a functional item, in terms of its capacities for mentation (i.e. engaging in mental activities). From a functionalist perspective, *what* makes a thing a thing is not what a thing is but what a thing does. In other words, the functional item is not independent of its activity.

The activities of the mind are indeed special in the sense that they are not ubiquitous. But as William Bechtel suggests it is not in spite of being comprised of mechanisms but *in virtue of* the right kind of mechanisms that the mind is special and its set of activities has distinctive characteristics.

For this reason, if the attack or the argument from the perspective of disabilities is adopted as a standard strategy toward machines or what Daniel Dennett calls "machine mentation" or if it is exercised as a pre-determined reaction to the possibility of the realization of the mind in different substrates, then it no longer enjoys a genuine critical attitude. Why? Because such a critical strategy then has implicitly subscribed itself to a preservationist view of the mind as something inherently foreclosed to mapping and (re)construction. The mind it safeguards has a special status because it is unique at the level of mapping and constructability. It cannot be constructed, because it cannot be fully mapped. It cannot be mapped because it cannot be defined. It cannot be defined because it is somewhere ineffable. If it is somewhere ineffable, then it is everywhere ineffable. Therefore, the singularity of the mind is the effect of its ineffability. If we buy into one ineffable thing and if that thing happens to be central to how we perceive the world, then we are also prepared to regard many other things in the universe as ineffable. Consequently, we have committed ourselves to full-blown mysticism.

Turing's AVD signals one of the most consequential phases in the historical development of the human and defining the project of humanity in the sense of both determining the meaning of being human and updating its definition. Its importance lies in how it grapples with the most fundamental question posed by Kant, "what is Man?", or what does it mean to be human?

Unlike the Copernican Revolution, the Darwinian Revolution, the Newtonian Revolution, and the Einsteinian Revolution in which we witness the consequences of the radical theoretical reorientation immediately manifesting themselves in the present world view, the site of the Turingian Revolution is always in the future. In short, the Turingian Revolution does not happen here and now, and for this reason, the conception of revolution it exercises fundamentally deviates from the trajectory of the Copernican Revolution.

The Turingian Revolution suggests that the future won't be a varied extension of the present condition. It won't be continuous to the present. Whatever arrives back from the future—which is in this case, both the mind implemented in a machine and a machine equipped with the mind—will be discontinuous to our historical anticipations regarding what the mind is and what the machine looks like.

But why is the Turingian Revolution in cognitive sciences and artificial intelligence the only revolution that is instantly conceived in and takes place in the future? Because what Turing proposes is a schema or a general program for a thoroughgoing reconstruction and revision of what it means to be human and by extension, the humanity as a collective and historical constellation. The underlying assumption of Turing is that the significance of human can be functionally abstracted and realized. This significance is the mind as a set of activities spanning perception, thinking, reasoning and the ability to engage in purposive action.

The adequate functional abstraction and realization of this account of human significance means 'what makes the human significant' can be realized by different individuating properties or realizers. But also what constitutes the human significance can be implemented in different modes of organization, material or otherwise. The contexts or the environments of the implementation have the ability to modify and update this functional schema drastically. In other words, the meaning of the mind will be changed through how it is used or how it is re-implemented, since implementation is not simply the relocation of a function or an abstract protocol from one supporting structure to another. It is the execution of a conceptual-functional schema in a new context and in a new environment with its specific sets of organizational demands. Accordingly, re-implementation is the repurposing and refashioning of a function that diversifies its pragmatic content.

Realizing the mind through the artificial by swapping its natural constitution or biological organization with other material or even social organizations is a central aspect of the mind. Being artificial, or more precisely, expressing itself via the artifactual is the very meaning of the mind as that which has a history rather than an essential nature. To have a history is to have the possibility of being artificial—that is to say, expressing yourself not by way of what is naturally given to you but by way of what you yourself can make and organize. Denouncing this history is the same as rejecting freedom in all its forms. Denying the artificial truth of the mind and refusing to take this truth to its ultimate conclusions is to antagonize the *history* of the mind, and therefore to be an enemy of thought.

The pragmatic functionalist understanding of the mind is a historical moment in the functional evolution of the mind. But evolution in what sense? In the sense that the pragmatic functionalist realization of the mind (the understanding of its meaning) coincides with the artificial realization of the mind (or the construction of its functional space by entirely different sets of realizers).

Once the real content of human significance is functionally abstracted, realized and implemented outside of its natural habitat, the link between the structure in which this function is embedded and the significance qua function is weakened. Up to now the influence of the structure (whether as a specific biological structure or a specific social stratum) over the function has been that of a constitution determining the behaviors or activities of the system. But with the abstraction and realization of those functions that distinguish the human—that is to say, by furnishing the real significance of the human with a functional autonomy—the link between the structure (or manifest humanity) and the function (all activities that make the human human) loses its determining power. The human significance qua function evolves despite the conditions under which it has been constituted.

If the determining influence of the constituting structure (in this case, the specific

biological substrate) over the function is sufficiently weakened, the image of the functional evolution can no longer be seen and recognized in the structure that supports it. The evolution at the level of function—here the expansion of the schema of the mind—is asymmetrical to the evolution of the structure, be it the evolution of the biological structure that once supported it or a new artificial habitat in which it is implemented. It is akin to a shadow that grows to the extent that it eclipses the body that once cast it.

In this fashion, what constituted or presently constitute the human no longer determines the consequences of what it means to be human. Why? Because, the functional realization of the meaning of being human implies the departure of this meaning from the present condition or the image with which we identify the human. To put it in more technical terms, the function is able to reconstitute itself by perpetually reconstructing and revising itself, by evolving asymmetrically with regard to the structure and by revising its meaning through re-implementation in new substrates. By being re-implemented, the function is able to change the schema of the mind. Turing's gesture toward the possibility of functional realization of the human—or more precisely, the activities which make the human human—means that human is always rewritten, redefined, reshaped from the future. If the functional realization of the mind suggests the modification and expansion of the schema of the mind irrespective of its present constitution (which is a basis for our manifest self-identification), then a program committed to the multiple realizability of human mind can no longer be characterized by recourse to the conditions of the past and the present.[1] It genuinely belongs to the future.

But the future—generally and specifically in terms of construction and revision—is discontinuous to the past and the present. Therefore, the constructive and revisionary dimension of Turing's functional realization of the human cannot be seen from the perspective of the present because the implications of construction and revision as the forces of reconstitution and reconception unfold from the future. In short, what Turing does is providing the blueprint of a program through which the consequences of being distinguished as human are discontinuous and irreconcilable with what we currently identify as the human.

Turning's functional realizability of human is a thesis about constructability. It suggests there is no essentialist limit to the reconstructability of the human or what really human significance consists in. However, it goes even further by proposing that the consequences of constructing the mind outside of its natural habitat, *reconstruction*

[1] According to the *multiple realizability thesis*, the realization of a function can be satisfied by different sets of realizing properties, individuating powers and activities. Therefore, the function can be realized in different environments outside of its natural habitat by different realizers. Multiple realizability usually comes in strong and constrained varieties. The strong version does not impose any material or organizational constraints on the realizability of a specific function, therefore the function is taken to be realizable in infinite ways or implementable in numerous substrates. The constrained variety, however, sees the conditions required for the realizability of a function through a deep or hierarchical model comprised of different explanatory levels and qualitatively different realizer properties which impose their respective constraints on the realization of the function. Each mechanism that explains an activity (qua explanandum) is a realist constraint upon that activity. But since mechanisms responsible for an activity are not uniformly or contiguously distributed, the constraints they impose are dimensionally varied. Accordingly, the criteria for the realization of a function are characterized as dimensionally varied and multiply constrained.

becomes tantamount to reconstitution. It is in this sense that Turing's project highlights a rupture in the truth of humanity, between the meaning of being human and its ramifications. It practically elaborates that to be human does not entail the understanding of the consequences of what it means to be human or coming to terms with such consequences. Indeed these two couldn't be further apart. To be human is neither sufficient condition for understanding what is happening to human nor is it a sufficient condition for recognizing what the human is becoming. It can neither fathom the consequences of revising the meaning of the human nor the scope of constructing the human according to this revisionary wave.

By functionally realizing the human, Turing draws a new link between emancipation (here the emancipation of human significance at the level of activities or functions) and the liberation of intelligence as a vector of self-realization. Both Turing's computationalism and functionalism are significant because the ramifications of these programs—no matter what their current state is and what setbacks they have suffered—cannot be thought by their present implications. In this sense, by definition humanity as we identify it in the present cannot grapple with and realize the scope of Turing's project.

In continuation of the project of the enlightenment, rather than being an argument for anti-humanism, Turing's Arguments from Various Disabilities (as appeared in his revolutionary 1950 essay *Computing Machinery and Intelligence*) is actually an argument for enlightened humanism insofar it fully conforms to the following principle: The consequentiality of the human and the human significance is not in its given meaning or a conserved and already decided definition, but in its ability to bootstrap complex abilities and intricate obligations out of primitive abilities and simple duties. These bootstrapping abilities that signify the human are precisely the expression of a fully updatable definition of the human as *a constructible and revisable mode of being*.

This is the underlying significance of Turing's project: The fact that the significance of the human lies not in its uniqueness or over-particularized mode of being but in its constructibility that allows for the updating of its definition, the upgrading of its abilities and its historical transformation from primitive forms to increasingly complex forms. This is nothing but the passage of Spirit as a functional constellation of rational agencies. In other words, Turing's computational project contributes to the project of enlightened humanism by dethroning human and ejecting it from the center while acknowledging the significance of the human in functionalist terms. For what is the expandable domain of computers if not the strongest assault upon the ratiocentricity of the human mind in favor of a view that the ratiocinating capacities of the human mind can be reconstructed and upgraded in the guise of machines?

It is the understanding of the meaning of human in functional terms that is but a blueprint for the reconstruction of human and the functional evolution of its significance beyond its present image. The knowledge of the mind as a functional entity develops into the exploration of possibilities of its reconstruction. While the exploration of functional realization by different realizers and for different purposes shapes the history of the mind as that which has no nature but only histories and possibilities of multiple realization.

What used to be called the human has now evolved beyond recognition. Narcissus can no longer see or anticipate his own image in the mirror. The recognition of the blank mirror is the sign that we have finally left our narcissistic phase behind. Indeed, we are undergoing a stage in which if humanity looks into the mirror it only sees an empty surface gawking back.