The Transhumanist Philosophy of Charles Sanders Peirce

Aaron B. Wilson
South Texas College
awilson3@southtexascollege.edu

Daniel J. Brunson
Morgan State University
daniel.brunson@morgan.edu

Abstract

We explain how the work of Charles Sanders Peirce (1839–1914) – the founder of semiotics and of the pragmatist tradition in philosophy – contributes an epistemological, metaphysical, and ethical foundation to some key transhumanist ideas, including the following claims: technological cognitive enhancement is not only possible but a present reality; pursuing more sweeping cognitive enhancements is epistemically rational; and current humans should try to evolve themselves into posthumans. On Peirce’s view, the fundamental aim of inquiry is truth, understood in terms of a stage of ideal cognition (what he calls the “final opinion”). As current human cognitive abilities are insufficient to achieve this stage, Peirce’s views on cognition support a variety of ways in which they might be enhanced. Finally, we argue that what Peirce describes as our ethical sumnum bonum seems remarkably similar to what Bostrom (2005) argues to be the core transhumanist value: “the exploration of the posthuman realm.”

Introduction

In “A History of Transhumanist Thought” (2005a), Nick Bostrom traces ideas about the promises of scientific progress (e.g. indefinite life extension, cognitive enhancement, and expanded control over nature) to a number of Renaissance and Enlightenment authors: Pico della Mirandola, Francis Bacon,
Offray de La Mettrie, and Nicolas de Condorcet. As influential as scientific humanism was during the early modern period, in the nineteenth century it strained against a tide of romantic thought and “latter-day reactions against the rule of instrumental reason in the attempt to rationally control nature” (Bostrom 2005a, 4) – which, in the works of Thoreau and others, often invoke primitivist or Luddite ideals. But nineteenth-century literature is far from bereft of transhumanist themes: some works by Mary Shelley, Jules Verne, and H.G. Wells are notable exceptions. And though Nietzsche’s concept of the Übermensch is not an expression of a technologically enhanced human, it is an expression of a person who, by rejecting all traditional gods, aspires to his own sort of godhood. The prominence of transhumanist thought grew throughout the twentieth century – with an explosion in the science fiction genre and with Julian Huxley (1951, 1957) seemingly coining the term “transhumanism” – before becoming a distinct intellectual movement and focus of academic study by the beginning of the twenty-first.

One late nineteenth- and early twentieth-century thinker whom Bostrom does not mention, but whom he’d have had good reason to include, is the scientist and philosopher Charles Sanders Peirce (1839–1914). Peirce is widely known as a founder of both modern semiotics (the theory of signs) and the pragmatist tradition in philosophy. But he should also be recognized for articulating ideas that, even if not (yet) particularly influential among transhumanist and posthumanist thinkers, can help guide transhumanist thought today. There are some suggestions in the relevant literature that Peirce may already have had some influence on transhumanism. For instance, in the “Philosophy of Transhumanism,” Max More mentions a “form of Piercean [sic] pragmatism” as a popular epistemological view among transhumanists (More 2013, 6). Peirce’s name does not come up often in the literature, but his philosophy offers a normative and metaphysical framework for transhumanism. Not only do Peirce’s writings on ultimate aims and values provide one possible basis for transhumanist aims and values, his accounts of cognition, inquiry, and truth entail both the possibility and the rationality of cognitive enhancement.

It is easy to miss this aspect of Peirce’s ideas, and Bostrom cannot be blamed for overlooking him. It is unclear whether Peirce had directly inspired any overtly transhumanist thinking at the time when Bostrom was writing. But any attentive reader who is willing to wade through the hundreds of pages of Peirce’s published and unpublished writings will notice that some elements of his philosophical system conspicuously align with transhumanism. One such element is his account of truth. Peirce identifies truth as both the ultimate aim of all inquiry and the ideal state of information. Assuming that human inquirers would need to enhance their cognitive abilities in order to indefinitely approach this ideal state, it becomes, on Peirce’s approach, epistemically or “logically” rational to do so. Further, Peirce upholds a form of the extended mind thesis, or the thesis that some of our own cognitive processes extend outside our skulls, which allows for much continuity between cognition and technology (e.g., our cognitive abilities may be extended into, and enhanced by, our smartphones). But perhaps the most fundamental connection between Peirce’s thought and transhumanism is with respect to values. This follows from his conception of the summum bonum, or the ultimate end of action, as well as a related conception of God. Peirce makes certain remarks suggesting that an approach to Godhood is our summum bonum (which comports with his view about truth). Bostrom (2005b) expresses a similar idea when he identifies the exploration of the posthuman realm as the ultimate aim or value for the transhumanist. So, notwithstanding some differences on details and terminology, there seems to be general agreement on core values between Peirce and Bostrom-style transhumanism.

We do not necessarily contend that Peirce, were he to have lived for another century, would have embraced the transhumanist label. We contend, rather, that certain of his philosophical views are well aligned with and support some key transhumanist ideas, so that it might reasonably be supposed that he would have embraced some current transhumanist thought (and added to it). In section 1., we explain how
the pursuit of artificial cognitive enhancement is required by Peirce’s accounts of truth and inquiry. In section 2, we examine Peirce’s metaphysics of cognition and how, based upon that metaphysics, our cognition can be directly enhanced by a wide range of technologies (including computers or “logical machines”). Finally, in section 3, we link several passages in Peirce’s work concerning God and the ethical *summum bonum* to show how the transformation of humans into posthumans would seem to be, on his view, a fundamental ethical priority.

1. Peircean truth and posthuman knowledge

Peirce’s identification of truth with “the opinion which is fated to be ultimately agreed to by all who investigate” (W3:237, 1878) – or with “the predestined result to which sufficient inquiry would ultimately lead” (CP 5.494, 1907) – is controversial. In response to the claim that there would be such a “final opinion” or final result of inquiry if it continued indefinitely, Bertrand Russell (1939, 146) asks a series of skeptical questions: “Is this an empirical generalization from the history of research? Or is it an optimistic belief in the perfectibility of man? Does it contain any element of prophecy, or is it a merely hypothetical statement of what would happen if men of science grew continually cleverer?” Another problem concerns how a “final opinion” could constitute the truth of a proposition; for even if a proposition’s being accepted as part of such an opinion is a sufficient condition for its truth, to many philosophers it could hardly be a necessary one.

Contrary to Russell, Peirce’s analysis of truth is not a “rash assertion” (1939, 146). Peirce has deep theoretical reasons for analyzing truth in terms of a final opinion, though his analysis does contain elements of empirical generalization and prophecy.

In his seminal 1878 *Popular Science Monthly* paper “How to Make Our Ideas Clear,” Peirce derives this analysis of truth by applying his rule for attaining the third (and highest) grade of clearness in our conceptions – what came to be called the “pragmatic maxim” – so that truth as the final opinion is supposed to be our clearest conception of truth. But it is mainly in writings of the 1900s that it becomes especially clear how the concept of the final opinion is supposed to be our clearest concept of truth. In these writings we find that, on the second grade of clearness, a proposition is true if and only if it corresponds to its object. But, Peirce asks, “what does this correspondence or reference of the sign, to its object, consist in?” (CP 5.553, 1906). His answer seems to be that it consists in a dynamical process that over time would tend toward a state of ideal stability or equilibrium between the interpreter of the sign and its object (this ideal state being “the final opinion”).

Though at some places Peirce writes as if there would be a distinct final opinion for every question (e.g, CP 8.43, 1885), he generally supposes that there would be one final opinion: the opinion that perfectly represents all reality external to the opinion or representation itself. This is apparent in his original 1878 paper. For reasons that we’ll return to, only as a whole could our beliefs attain an ideal stability with external reality. Moreover, while Peirce holds that our clearest conception of reality is as the object of the final opinion, which entails that there is no reality other than that which would be represented in the final opinion, this does not entail that only at the final opinion is anything real. On its second grade of clearness, reality is “that whose characters are independent of what anyone thinks them to be” (CP 5.405, 1878); and, on the second grade, truth is a certain correspondence between a representation and something that exists independently of any representation of it.

But it is precisely this idea of reality as mind-independent that leads Peirce to the analysis of truth as the final opinion. In “Some Consequences of Four Incapacities” (1868), he argues that the mind-
independence conception arises in us when “we discover that there was an unreal, an illusion; that is, when we first corrected ourselves” (W3:239). We discover that reality is not simply whatever we might expect or want it to be, but is, instead, a vast “non-ego” that incessantly clashes with our desires and expectations. When reality disappoints our expectations, often we suppose that we made some error and we attempt to correct ourselves (that is, if we are not in a state of denial). Peirce calls the deliberate attempt to self-correct inquiry; and, on his view, the mind is essentially a self-correcting system. What he calls “the law of mind” is a tendency to develop habits of thought and belief that are better adapted to achieving a certain end—namely, the avoidance of disappointment by means of the use of things as signs. The theoretical final result of this tendency would be an ideal stability between our use of signs and our perceptual and behavioral interactions with objects. At that stage, there would no longer be any need to self-correct. It is precisely in this sense that we would achieve a sort of “complete knowledge” of external reality. There would be nothing left for mind to know except itself (for each representation of itself would introduce a new representation to represent the previous one, ad infinitum). 7

Whether or not we make a deliberate attempt to self-correct, our inherent disposition to self-correct would eventually compel us to believe the truth. Peirce writes that even “the most pigheaded and passionate of men who has sworn by all the gods that he never will allow himself to believe the earth is round” will “surely come to and rest in the truth about the form of the earth”—that is, if we “give him time enough, and cram that time with experience in the pertinent sphere” (CP 7.78, c.1905). But so long as something remains unknown, the process of self-correction continues. Indeed, knowing that there may remain something unknown can cause doubt, as we might not know whether that unknown might affect the justification of a given belief. Thus, any belief fails to achieve ideal equilibrium with external reality if some part of it remains unknown. The truth of any belief, then, depends on its being held alongside other true beliefs; and the pursuit of the truth on any question is the pursuit of all truth—the pursuit of the final opinion. 8

Peirce’s account of the deliberate level of the self-corrective process—i.e. inquiry—makes the most direct contribution to transhumanist thinking. While in his 1877–78 papers he argues that the goal of any sincere inquiry is truth, 9 in works of the 1900s Peirce explicates his general normative theory according to which truth is the ultimate aim of inquiry, such that the adoption of truth as its aim is constitutive of any genuine inquiry. Since the ideal stability of belief is achieved only in toto, then the goal of any inquiry can be none other than the final opinion. The final opinion is what all rational inquirers must aim to achieve.

On Peirce’s account, then, not only is it logically rational to pursue this stage of perfect cognition or ideally stable belief, it is logically irrational not to pursue it (though under many circumstances it might not be ethical to pursue it). Here a quintessential transhumanist idea emerges from Peirce: the idea that we ought to try to achieve a posthuman stage of knowledge. As a perfect stage of knowledge, the final opinion may also be described as a posthuman stage of knowledge—that is, on the assumption that humans in their current form are incapable of indefinitely approaching the final opinion. There are some indications that Peirce himself held this assumption. For example, he consistently argues that science thus far is little more than the development of instincts limited by the needs of evolution (e.g., CP 1.118, 5.586, 6.491, 7.508). Indeed, he suggests that all knowledge possessed by current humans is built upon biologically based instincts that will need to cease “as we penetrate further and further from the surface of nature” (CP 7.508, c.1898). Peirce thus expresses the idea that humans may eventually need to change fundamentally in order to advance any further in their knowledge.

Nonetheless, it is irrelevant whether Peirce believed that humans are incapable of understanding nature outside of the “moderate-sized specimens of dry goods” of their ecological niche (Austin 1964, 8). On his
theory, it is true that if current humans are incapable of indefinitely approaching the final opinion, then it is logically rational for them to pursue the enhancements by which they could become so capable. Transhumanists generally assume that there are some realities that are inaccessible to humans given their biological limitations; and this assumption is both intuitive and empirically supportable.

The fact that there is so much we know to be cognitively inaccessible to other animals is often used to support the claim that there is probably much that is inaccessible to current humans. Also, there seem to be realities open to animals that are not open to humans – at the least, facts about “what it is like” to be other animals (i.e., facts about their subjective experience). Indeed, even if basic physics is nearing “completion” (a proposition many view as highly doubtful), there still may be facts about subjective experiences that we cannot represent in our current biological form. Bostrom assumes at least this much, but he thinks it is safe to assume that there are realities accessible only to posthumans: creatures with cognitive and sensory powers that far surpass those of any human. He offers the following diagram to illustrate this idea:

![Diagram of the Space of Possible Modes of Being]

It is possible that there are realities inaccessible even to posthumans, and Bostrom represents that in this diagram. Facts about the subjective experience of animals, including current humans, might be among them. Although Peirce holds that all reality would be accessible to the being or beings that achieve the final opinion, he can allow that at least current humans are limited in terms of the realities that are accessible to them.

Upon the assumption that some realities are inaccessible to current humans, a simple argument supporting the creation of superhuman cognition emerges from Peirce’s theory of truth:

1. _At the final opinion, everything real is represented._

2. _It is likely that reality exceeds what is within the cognitive grasp of humans given our current biological limitations._
3. Thus, it is likely that the final opinion can be indefinitely approached only by some intelligence surpassing current human limitations.

Not only is this argument implied by Peirce’s descriptive and normative accounts of truth (in connection with his belief that reality may far exceed what we can currently fathom), it also accords with his own optimism about the promises of technology, including its promises of enhancing our abilities to acquire knowledge.

Peirce’s optimism about what can be known (given enough time) shows clearly in his response to “the problem of buried secrets.” This is the problem that some “minute facts of history,” such as the precise history of the first self-replicating molecules on Earth, seem impossible to know. His response takes the form a rhetorical question: if inquiry “were to go on for a million, or a billion, or any number of years you please, how is it possible to say that there is any question which might not ultimately be solved” – especially, he says, “given the activity of the last hundred?” (CP 5.409, 1878). More specially, he opines that the potential of science is limitless: “Give science only a hundred more centuries of increase in geometrical progression, and she may be expected to find that the sound waves of Aristotle’s voice have somehow recorded themselves” (CP 5.542, 1902). Peirce does not bet that any question will be answered, as a matter of necessity. He asserts only that any question can and would be answered, so long as it is coherent and given enough time and experience.

Further, Peirce was deeply impressed with the technological advancements achieved by his time: “Modern science, with its microscopes and its telescopes, with its chemistry and electricity, and with its entirely new appliances of life, has put us into quite another world; almost as much so as if it had transported our race to another planet” (CP 5.513). He recognizes and values how instruments such as microscopes and telescopes afford us “recondite experiences” (CP 8.110, 1900) that advance us toward truth. Familiar with the work of Charles Babbage (1791–1871) and others (see CP 2.56, 1902 and 4.611, 1908), Peirce understood the potential of calculating machines to perform logical inferences, and he may have been the first to recognize the potential of machines literally to think. In his 1887 article “Logical Machines,” he comments: “Precisely how much of the business of thinking a machine could possibly be made to perform, and what part of it must be left to a living mind, is a question not without conceivable practical importance” (NEM 3: 625–32). In 1902, he considers the possibility that such instruments could even have souls, given that they could think (CP 2.56). Though he concludes that logical machines do not have souls – at least the logical machines conceivable in his time – he clearly considers whether they do to be a reasonable question.

It is hard to imagine, then, that Peirce would not see it as logically rational for us to modify ourselves through applied science or to engineer some form of superintelligence, if doing so is necessary for the indefinite community of inquirers to continue to advance toward the final opinion. Since truth as the final opinion is the constitutive aim of any genuine inquiry, it is thus instrumentally rational for inquirers to enhance themselves artificially, if doing so is necessary or is simply the most expeditious way to advance. Thus, combining this with the notion that some realities are inaccessible to current humans, we obtain the following argument:

1. All reality is knowable only through cognitive abilities that exceed those of current humans.

2. If all reality is knowable only through superhuman cognitive abilities, then we ought to try to engineer those abilities.
3. *Therefore, we ought to try to engineer superhuman cognitive abilities.*

This argument also implies that it might be rational to create a form of superintelligence that is *not* continuous with current humans, or that does not emerge from direct enhancement of current humans. Peirce views the approach toward the final opinion as contingent upon, not the survival of the human race in particular, but the survival of the “indefinite community” of inquirers (e.g., W2:271, 1869; CP 2.654, 1878). He declares that “interest in an indefinite community” and “hope in the unlimited continuance of intellectual activity” are “indispensable requirements of logic” (CP 2.655).^{15}

But if the enhancement of current humans is a means of advancing inquiry generally, we should aim to develop the technologies by which we could enhance our own cognitive abilities. On Peirce’s view, by the first years of the twentieth century, we had already developed technologies by which our minds are not only enhanced, but are also *externalized* – in that parts of our cognitive processes take place outside our skulls.

2. Peirce’s external mind

The *locus classicus* for contemporary discussions of the “extended mind” remains Clark and Chalmers’ 1998 argument for *active externalism*. They argue against the “hegemony of skin and skull” in favor of cognitive agents that are not only embodied, but also *embedded* within external contexts that, at minimum, support cognition. In a riff on another pragmatist, Hilary Putnam, Clark and Chalmers summarize their argument as follows:

> If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain’t (all) in the head! (1998, 8)

This view is called “active externalism” because it goes beyond externalism regarding mental contents, which Clark and Chalmers see as having only a passive effect on cognition, and instead makes central a coupling of organism and environment “that can be seen as a cognitive system in its own right” (1998, 8). To make this plausible, they introduce a thought experiment involving Otto, a person with Alzheimer’s disease who relies upon a notebook to supplement, or even replace, his memory. Without his notebook, he is prone to cognitive failures, but with it he functions as well as a “normal” cognitive agent. This suggests that Otto’s very self is extended: “The information in Otto’s notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that Otto *himself* is best regarded as an extended system, a coupling of biological organism and external resources” (1998, 18).

What is true of Otto is true of many of us, especially in light of the proliferation of cognitive prostheses – i.e., smartphones – since Clark and Chalmers’ original paper. While our focus is on the technological extension and improvement of the human central to transhumanism, we also want to note that the “external resources” mentioned by Clark and Chalmers sometimes include other organisms:

> What about socially extended cognition? Could my mental states be constituted by the states of other thinkers? We see no reason why not, in principle. In an unusually interdependent couple, it is entirely possible that one partner’s beliefs will play the same sort of role for the other as the notebook plays for Otto. (1998, 17)
However, why think this is limited to “unusually” interdependent couples? As Daniel Wegner has argued, we are enmeshed in a variety of transactive memory systems that, while perhaps idealized by a couple that literally finishes each other’s thoughts, may involve looser social ties, such as between professor and student. Wegner (1986) also acknowledges that artifacts contribute to our transactive memory systems, insofar as we delegate much of our memory storage to external locations:

As an example of this, suppose that one encodes this book in memory. To encode it internally requires learning some sort of label (e.g., the book’s title), and then memorizing the rest of it from cover to cover. To encode it externally also requires learning the label, but then only calls for learning where the book can be found. (Wegner 1986, 189)

Thus, whether artifact or person, our cognition reaches into the world.

Significantly, Peirce anticipated these forms of active and transactive externalism, and thereby provided a potential transhumanist ontology in addition to an epistemology and ethics. In 1868’s “Some Consequences of Four Incapacities” and elsewhere, Peirce argues for “scholastic realism”, the position that generals or universals (e.g., red, man) and not only particulars (e.g., this red, Aristotle) have a real existence (CP 5.312). This realism leads to the position that “the mind is a sign developing according to the laws of inference” and therefore, perhaps bizarrely, there are only relative differences between a person and a word (CP 5.313). In short, people and words are both signs, distinct from their physical manifestations, whether my body before lunch and after, or in thousands of different books:

It is hard for man to understand this, because he persists in identifying himself with his will, his power over the animal organism, with brute force. Now the organism is only an instrument of thought. But the identity of a man consists in the consistency of what he does and thinks, and consistency is the intellectual character of a thing; that is, is its expressing something. (CP 5.315)

This leads Peirce to claim nearly 40 years later, in 1905’s “What Pragmatism Is,” that “the man’s circle of society (however widely or narrowly this phrase may be understood), is sort of a loosely compacted person, in some respects of higher rank than the person of an individual organism” (CP 5.421). Thus, Peirce seems to be a clear predecessor to the tradition of socially extended transactive cognition of Clark, Chalmers, and Wegner, among others. He even has his own take on the “hegemony of skull and skin,” as evinced by the following:

Why we used to read that the soul resides in a little organ of the brain no bigger than a pin’s head. Most anthropologists now more rationally say that the soul is either spread over the whole body or is all in all and all in every part. But are we shut up in a box of flesh and blood? When I communicate my thought and my sentiments to a friend with whom I am in full sympathy, so that my feelings pass into him and I am conscious of what he feels, do I not live in his brain as well as in my own – most literally? (CP. 7.591; emphasis added)

Peirce goes further to argue that cognition is extended not only socially, but also artifactually and technologically. He remarks that “it is much more true that the thoughts of a living writer are in any printed copy of his book than that they are in his brain.” This comes within the context of a distinction he makes between mind and consciousness, where consciousness “is nothing but the inward aspect of things, while mind on the contrary is essentially an external phenomenon” (CP 7.364, 1902). By “external” he does not mean only “external to the skull,” but he is clear that mind can be partly external to the skull.
This is shown not only by his comments about an author’s thoughts being more in his book than in his brain, but also by his general argument concerning the localization of the mind:

A psychologist cuts out a lobe of my brain and then, when I find I cannot express myself, he says, “You see, your faculty of language was localized in that lobe.” No doubt it was; and so, if he had filched my inkstand, I should not have been able to continue my discussion until I had got another. Yea, the very thoughts would not come to me. So my faculty of discussion is equally localized in my inkstand. (CP 7.366, 1905)

Peirce insists that the inkstand is not only a means of communicating his thoughts, but is also an engine or co-producer of his thoughts. As Skagestad puts it, “thoughts come to him in and through the act of writing, so that having writing implements is a condition for having certain thoughts” (1999, 551). Though a condition for thought is not the same as an embodiment or a localization of thought, Peirce’s point is precisely that if our best reason for locating the mind in the brain is that brain processes and organs are necessary for thought, then we have the same reason to locate the mind outside of the brain, in objects like books, inkstands, and, in the early twenty-first century, in smartphones, personal computers and the World Wide Web. Finally, Peirce makes an intriguing suggestion in one of his accounts of cognition as a process of observation and experimentation “which is substantially the same process whether it be performed with physical apparatus such as the chemist uses or with an apparatus of diagrams of our own creation, such as the mathematician employ” (CP 7.276). These “diagrams of our own creation” can be written equations or figures, or simply imagined ones. Thus, for Peirce, we can reason through a physical apparatus, such as a chemistry set, in effectively the same sense in which we reason through our own brains.

Peirce’s own extended mind thesis thus allows for more forms of cognitive enhancement than the old localized(-in-the-brain) view. Our increasing ability to interface our brains with our computers will almost surely enhance our ability to perform many different cognitive tasks. We already interface our brains with our computers using our eyes and our hands, and this gives us vastly more powerful ways to enhance our ability to perform mathematical, linguistic, and knowledge-retrieval tasks. Further, Peirce argues that the growth of science is due, in large part, to the development of technologies to expand our biological faculties. He himself worked for the U.S. Coast Geodetic Survey, and developed an improved pendulum for measuring minute variations in gravity (CP 7.16). In the century of change after his death, we have already moved far beyond the microscopes, pendulums, and telescopes that Peirce thought had put us into another world of ability and knowledge.

However, why do we necessarily want to enhance our abilities? On Peirce’s view, it is not simply to expedite our approach to the final opinion. It is also to expedite our approach to a state that we might call posthumanity.

3. Peirce’s summum bonum and the posthuman realm

In section 1., we argued that, upon Peirce’s account, it is logically or epistemically rational for human inquirers to enhance their cognitive abilities through technological means in order to surpass biological barriers to approaching the final opinion or truth. But while it is logically or epistemically rational to do so, there are many circumstances under which it would not be practical, morally correct, or rational in an all-things-considered sense, to enhance or to promote such enhancement. For instance, the development of certain relevant technologies may pose a significant risk to the welfare of millions. Many transhumanist thinkers, notably Bostrom (2014), have investigated and warned of such risks. But Peirce’s
moral philosophy suggests that, while these risks are morally serious, we have a fundamental moral obligation to pursue enhancements of our cognitive, social, and physical capacities. On his view, it seems, our fundamental moral imperative is to strive toward a sort of Godhood, where this imperative includes our fundamental epistemic imperative to strive toward a Godlike stage of knowledge (i.e., the final opinion).

Here there is a distinct correspondence between Peircean ethics and Bostrom-style transhumanism. Peirce’s *sumnum bonum* bears striking similarity to what Bostrom (2005b) declares to be the core transhumanist ideal: *the exploration of the posthuman realm*. At some point in their approach toward the final opinion, enhanced human inquirers will have enhanced so far that little or no human biology will remain essential to their continued existence. They will have become posthumans. Thus, approaching the final opinion implies approaching what Bostrom calls the posthuman realm.

By the “posthuman realm” Bostrom means the “modes of being” occupied by creatures that have physical, cognitive, and moral capacities that far exceed those of current humans. Potential posthumans include descendants of humans and, most optimistically, former humans who have attained such extraordinary capacities over time. He does not assume that there is any sharp transition from human to transhuman (or enhanced human) to posthuman: a posthuman may just be a human that is so enhanced, perhaps shedding all of its original biology, that it becomes unclear what meaning there is in calling it “human.” The *exploration of the posthuman realm*, then, would be the exploration of what is accessible with such capacities. It would be precipitous to place metaphysical limitations on this exploration, other than that it must involve the exploration of more than what is accessible to current humans. But for Bostrom, it potentially involves the exploration of other values, as current human values might not be among the values of posthumans. Our current values would be more deeply explored, as Bostrom points out, but no specific values or motivations can be established a priori for posthumans.

Peirce always considered himself first and foremost a logician. But at least from the late 1870s onward, ethical matters were never very far from his mind. He says that he began taking a serious interest in moral philosophy in 1883 (CP 5.129), which was around the time he wrote a review of Josiah Royce’s *The Religious Aspect of Philosophy* (1885), in which he criticizes Royce’s distinction between “moral realism” and “moral idealism” and seems to defend a sort of synthesis of the two (CP 8.45–52). While in his famous 1898 Cambridge Conferences lectures, Peirce remarks that ethics is a “useless” science, by 1902 he begins to place far more systematic importance on ethics. He declares that ethics is a positive science or a science concerning real facts, and he develops a general value theory on which the *sumnum bonum* or the highest ideal of human action incorporates the highest logical ideal: the final opinion.

In his 1903 Harvard lectures (and in other writings around that time), Peirce identifies logic as a normative science, where normative science generally concerns “the laws of the relation of phenomena to ends” (EP2:197). As such, logic concerns the relation of phenomena to truth – the ultimate end of inquiry (EP 2:200). Peirce’s use of “logic” as the name for an area of inquiry includes what (within the last century of mostly Anglophone philosophy) has gone under the headings of “epistemology” and “philosophy of science,” as well as “philosophy of language.” Peirce also considers ethics a normative science, where ethics concerns “those things whose end lie in action” (EP 2:200). Ethics “is the study of what ends of action we are deliberately prepared to adopt” – or of what is “morally good” (CP 5.130). It is not only the study of the conformity of action to an end or ideal, but is the “theory of the ideal itself, the nature of the *sumnum bonum*” (EP2:377, 1906). Note, however, that Peirce thinks that “the morally good appears as a particular species of the esthetically good” (CP 5.130), because the moral good must be “admirable in itself” and esthetics is the study of what is admirable in itself. And just as the moral good
(the *summum bonum*) is a species of the esthetic good, so the logical good (truth) is, on his view, a species of the moral good (CP 5.130). He writes that truth is a “phase” of the *summum bonum*; and for this reason “the logician ought to recognize what our ultimate aim [of action] is” (CP 1.611, 1903).

But what exactly does Peirce take to be the *summum bonum*? What he says about it is rather cryptic and a little sparse. He writes that, for the pragmaticist, the *summum bonum* is the “process of evolution whereby the existent comes more and more to embody [reasonable] generals,” where at higher stages this “evolution takes place more and more largely through self-control” (CP 5.433, 1905). Many scholars identify this “process of evolution” with what he elsewhere (and a little earlier) refers to as the development of “concrete reasonableness” (CP 5.3, 1902; see Potter 1967, Herdy 2014, Atkins 2016), which involves “the becoming governed by laws” – that is, bringing order to chaos. Laws are the generals which the existent grows to embody; and while, at earlier stages, the formation of laws is not necessarily guided by any intelligence (as is the case with basic physical laws), at later stages the formation of laws occurs through intelligent self-controlled processes (as is the case with social or institutional laws – this is not to imply that all institutional laws should be thought of as intelligent!). But the process cannot end with the formation of institutional laws, for there is far more room for growth of order in the universe. Peirce’s idea seems to be that the development of order across *all* reality would become guided by intelligent processes. This implies an intelligent creator – or, perhaps, an intelligent *re*-creator.

In 1905, Peirce writes that “I look upon creation as going on and I believe that such vague ideas as we can have of the power of creation is best identified with the idea of *theism*” (CP 8.138fn4, 1905; our emphasis). Continuing in the same note, he tells us that “the [ultimate] ideal would be to be fulfilling our appropriate offices in the work of creation,” where “every man” should “feel that he is doing what God made him in order that he should do”. Peirce seems committed to the idea of a pre-creation God, as opposed to a created God that takes over the process of creation. And there are clear grounds for regarding him as a traditional theist (for more on Peirce’s complex conception of God, see Delaney 1992, Mahowald 1976, and Orange 1985).

However, Peirce complicates his theism at several points in his writings. A stunning example is the following passage, where he cautions against claiming that *God exists* in any “wooden sense” and hints at a God that “self-reproduces”:

> We cannot possibly put ourselves in God’s shoes, even so far as to say in any definite wooden sense that *God is*. I only mean that the purpose of creation as it must appear to us in our highest approaches to an understanding of it, is to make an answering mind. It is God’s movement toward self-reproduction. And when I say that *God is*, I mean that the conception of a God is the highest flight toward an understanding of the original of the whole physico-psychical universe that we can make. (R 1334, 1905)

Peirce says that the “purpose of creation” is to make an “answering mind”, and that to make an answering mind is “God’s movement toward self-reproduction.” The creation of an “answering mind” would seem to be the achievement of a mind that answers all questions or achieves the final opinion (as one “perfection” among others). It seems that, based on this passage, God’s purpose behind creation is to produce a new God, or perhaps to “renew Himself.” If this is supposed to occur through the self-controlled and intelligent conduct of created beings, then the *creation of a new God* is the *summum bonum* of intelligent created beings like us.
Moreover, Peirce suggests that we should not necessarily look to God for help in pursuit of this end. We should pursue it on our own. He remarks that “[l]ike a good teacher, [God] is engaged in detaching us from a false dependence upon him” (CP 6.507). At some places Peirce speaks of a “love of God”; and in the 1885 review of Royce, he identifies love of God with love of an “ideal and divine humanity” (CP 8.47). He argues that this love develops from a “collective egoism” (an in-group bias) and grows into a “sympathy” for the “ocean of human misery” before “finally, steadying itself by the conception of an ideal humanity and a divine providence” (CP 8.49). If this love of “an ideal and divine humanity” develops from a sympathy for all human suffering, presumably it is a love, or a hope, for a humanity that will become ideal and divine, and does not suffer. It would seem, then, that this love for God is essentially a desire for humans to overcome those conditions that limit their freedom and cause them suffering.

To be logical, Peirce writes, “men should not be selfish” (CP 2.654, 1878); and this is because the “ideal perfection of knowledge” is the goal, not of an individual, but of the entire “community of inquirers” (W2:271, 1869). We should not desire that just humans approach the final opinion. As Peirce says, we must extend this community of inquirers to “all races of beings with whom we can come in immediate or mediate intellectual relation” (CP 2.654, 1878). And so, as our ultimate logical end, truth or the final opinion, is a phase of the sumnum bonum, the pursuit of it implies the pursuit of an ideal and divine community – one that is not limited by natural biological conditions.

Peirce thus anticipates Hopkins (2008), who writes: “these elements of a transhumanist moral vision – namely that we alter ourselves in order to pursue the humanly imagined and unimagined good that we cannot achieve given our current limitations, only speaks to the grander picture, the ultimate motivation, the search for the sumnum bonum of the fully sapient creature” (Hopkins 2008, 6). Of course, Peirce never says explicitly that we should enhance our basic capacities through technological means – a task that was practically unheard of during his time. While he attended to the ways we already think in and through our environment, including artifacts, he did not call for the invention of better inkstands or typewriters. Nevertheless, the idea that we should enhance our basic capacities is strongly implied by his normative philosophy, or by his conceptions of the ultimate end of all inquiry, and, more generally, the ultimate end of all conduct.

One of the central points of Bostrom’s work on transhumanist values is that the posthuman might explore values that are inaccessible to current humans for one reason or another. We find a not-so-dissimilar idea in Peirce. The sumnum bonum – the growth of concrete reasonableness through a self-controlled approach toward a God-like stage of being – is the moral good for current humans; but the moral good is, on his view, a species of the esthetically good, which is anything that is admirable-in-itself, including things that are not directly goods of action. “Beauty” might be an example of this, as beauty is not (or is not strictly) a good of action. Indeed, the approach toward Godhood or toward the posthuman may be the sumnum bonum because it involves the realization of all that is inherently admirable or valuable.

Peirce offers a consistent vision of the human, a natural being, both limited by bodies and already transcending them, on a trajectory of communal inquiry that is also a project of self-overcoming in order to know all that is knowable, and to experience all that can be experienced. Thus, Peirce deserves to be regarded as one of the great philosophical predecessors to contemporary transhumanism.
Notes

1. We would like to thank the participants of “Science and Values in Peirce and Dewey: A Conference in Honour of Angus Kerr-Lawson” at the University of Waterloo, especially the organizer, Shannon Dea, for an initial opportunity to present these ideas.

2. Bostrom also traces transhumanist thought to some ancient schools, myths, and literature, such as the *Epic of Gilgamesh* (c. 2000 B.C.).

3. As the name of a distinct intellectual movement, the term “transhumanism” originates with Max More and was first used by a circle of associates in California that also included FM-2030 (originally Fereidoun M. Esfandiary), and Natasha Vita-More. Although, in More’s original publication, “transhumanism” is defined as “a class of philosophies that seek to guide us towards a posthuman condition” (1990, 6), Julian Huxley had previously characterized transhumanism as the belief that “the human species can, if it wishes, transcend itself not just sporadically, an individual here in one way, an individual there in another way, but in its entirety, as humanity” (1957, 17). More does not appear to have borrowed the term from Huxley. Nonetheless, the latter’s conception of transhumanism would seem to fall within the class of philosophies that More calls “transhumanism.”

4. We have found no evidence that Peirce had a direct influence on the origins of transhumanism, including the term. However, Harrison and Wolyniak 2015 argue, *contra* Bostrom, that Huxley did not coin the term. Instead, they find it in Dante’s *Paradiso*, which seemed to influence a 1940 reference to “Paul’s Transhumanism” by W. D. Lighthall in his “The Law of Cosmic Evolutionary Adaptation.” Lighthall was also influenced by a fellow Canadian philosopher, Bruce W. Brotherston, referencing him in a 1926 article on “superpersonality.” Brotherston knew the classical pragmatists, and even published an article on Peirce in 1939. However, the trail turns cold here with the slimmest of possible influence. Indirectly, Peirce’s potential influence via semiotics (especially cybersemiotics) is too immense to review here; see Brier 1999 for one example.

5. As used here, “epistemic rationality” applies to deliberate actions that would maximize the realization of epistemic ideals such as the attainment of truth, either by increasing the likelihood of achieving an epistemic ideal, by expediting the time it would take to achieve an epistemic ideal, or by realizing something that approximates closer to an epistemic ideal. In the context of the work of Charles S. Peirce, “logical rationality” will be used as a synonym. On his usage, “logic” concerns representation, reasoning, inquiry, and knowledge. For a full account of Peirce’s use of “logic,” see Wilson 2016, 34–35.

6. In 1903, Peirce wrote: “Every man is fully satisfied that there is such a thing as truth, or he would not ask any question. *That* truth consists in a conformity to something independent of his thinking it to be so, or of any man’s opinion on that subject” (CP 5.211; Peirce’s emphasis).

7. As Wilson (2016) explains,

   a complete representation of reality, then, would really represent every real representation. This leads to a regress: any representation A is represented by B, which in turn is represented by C, ad infinitum. … But this infinite regress occurs only with respect to self-referential knowledge or representation, which leaves room for the possibility of an actual state at which all that remains to be represented are the representations of the final opinion itself. (256–57)
Wilson points out other possible problems with conceiving Peirce’s “final opinion” as an achievable state, as opposed to be a purely theoretical (unachievable) limit. But so long as reality external to the representational process itself is finite, Peirce’s “final opinion” is an achievable point. If external reality is not finite, then the semiotic process cannot be described as tending toward even a limit.

8. That is, the pursuit of truth is the pursuit of “all truth” about external reality; see the previous note.

9. In the first of his 1877–1878 Popular Science Monthly series of articles, “The Fixation of Belief,” Peirce argues that the “settlement of opinion” is the sole aim of any inquiry (W 3:248), and in the second of the series, “How to Make Our Ideas Clear,” he identifies the “opinion which is fated to be ultimately agreed to by all who investigate” as our clearest conception of truth.

10. The belief that there are such facts about subjective experience springs from such famous “knowledge arguments” as those given by Nagel 1974 and Jackson 1982. Jackson’s famous argument involves a scientist who never sees colors but learns all there is to know about the physical universe from a black and white screen in a black and white room. Upon being released, she sees colors, and the intuition is that she learns something new. Some maintain that she learns something new, but not a different type of fact: a different “way” of knowing a certain fact (Horgan 1984); others maintain that she only acquires new non-propositional knowledge, acquaintance-knowledge and knowledge-how, and that therefore she is not learning any new facts (Lewis 1988; Conee 1994).


12. There are at least two other passages in which Peirce argues that our science is limited by our natural instincts: (1) “In the first place all that science has done is to study those relations between objects which were brought into prominence and conceiving which we had been endowed with some original knowledge in two instincts – the instinct of feeding, which brought with it elementary knowledge of mechanical forces, space, etc., and the instinct of breeding, which brought with it elementary knowledge of psychical motives, of time, etc.” (CP 1.118, c.1896); and (2) “The instincts connected with the need of nutrition have furnished all animals with some virtual knowledge of space and of force, and made them applied physicists. The instincts connected with sexual reproduction have furnished all animals at all like ourselves with some virtual comprehension of the minds of other animals of their kind, so that they are applied psychists. Now not only our accomplished science, but even our scientific questions have been pretty exclusively limited to the development of those two branches of natural knowledge” (CP 5.586, 1898).

13. More specifically, at the final opinion, everything that is real and external to the representation (the final opinion) itself is represented. See note 7.

14. Also see CP 2.642, 1878, where Peirce suggests the possibility of learning more about Napoleon Bonaparte:

It may be that, at the time of his career, events were being recorded in some way not now dreamed of, that some ingenious creature on a neighboring planet was photographing the earth, and that these pictures on a sufficiently large scale may some time come into our possession, or that some mirror upon a distant star will, when the light reaches it, reflect the whole story back to earth. Never mind how improbable these suppositions are; everything which happens is infinitely improbable.
15. Peirce argues:

Our perversity and that of others may indefinitely postpone the settlement of opinion; it might even conceivably cause an arbitrary proposition to be universally accepted as long as the human race should last. Yet even that would not change the nature of the belief, which alone could be the result of investigation carried sufficiently far; and if, after the extinction of our race, another should arise with faculties and disposition for investigation, that true opinion must be the one which they would ultimately come to. (CP 5.408, 1878)

16. See Wegner 1986, 189:

In either case, one person has access to information in another’s memory by virtue of knowing that the other person is a location for an item with a certain label. This allows both people to depend on communication with each other for the enhancement of their personal memory stores. At the same time, however, this interdependence produces a knowledge-holding system that is larger and more complex than either of the individuals’ own memory systems.

17. In later works, Peirce is more careful to distinguish between existence and reality as modes of being, while still holding that universals are real. For more on memory and Peirce’s pragmatism and realism, see Brunson 2007.

18. As “soul” is not a technical term for Peirce, we could readily replace it with “mind” in this passage. In addition, there are some suggestions in Peirce (e.g., CP 5.313, 7.592) that could support reinterpreting his views along the lines of Floridi’s contention that personal identity is inherently informational – see Floridi 2011, and also Fairbanks 1976. However, we cannot develop this argument here.

19. As Esposito argues: “In general, the newer sciences of cybernetics, bionics, and ergonomics can perhaps give new vitality to [Peirce’s] synechistic philosophy by refusing to take seriously, whether consciously or by implication, essential mind/matter and natural/artificial distinctions” (1973, 76). *Synechism* is Peirce’s name for “the tendency to regard everything as continuous” (CP 7.565).

20. There may be physical and even psychological continuity between a human and a posthuman. But the differences between a human and a posthuman may exceed the differences between a bacterium and a human; in that case, a whole new classification of lifeform (a kingdom) would be appropriate. Indeed, the concept of the posthuman challenges our very definition of “life.”

21. One might attempt to place some metaphysical limitations on the posthuman realm through a priori reasoning (e.g., through application of the principle of non-contradiction). But a fundamental assumption of (Peircean) fallibilism is that anything is epistemically possible: for any proposition \( P \) that S thinks that she knows, it is possible that S does not know that \( P \) (because not-\( P \)). This is different from skepticism, which says that for any \( P \) that S thinks that she knows, S does not know that \( P \). (Peircean) fallibilism does not rule out that S knows that \( P \), or even that S knows that she knows that \( P \). It is the claim that any knowledge-claim could be false, so that, for us, anything could be true or false in that sense. Posthumans might be in a totally different epistemic situation.

22. For instance, Peirce identifies logic as a normative science in the 1902 unpublished fragments of the *Minute Logic* (e.g. CP 1.281, 1.575).
23. Peirce adds that it is the business of “the moralist” to find out what our ultimate aim is; but whatever the moralist finds it to be, “the logician has to accept the teaching of ethics in this regard” (CP 1.611, 1903).

24. As is often noted in works on Peirce, in 1905 Peirce adopted the term “pragmaticism” to distinguish his form of pragmatism from others. By this point in his writings, his pragmatism or pragmaticism was clearly a theory of conceptual representation – and not just a principle of analysis – of which he took his realism about modality and universal as well as his “critical common-sensism” or his position on first principles to be consequences (CP 5.439).

25. Perhaps the closest he gets is the practical recommendation to constantly take notes:

    Moreover, procure, in lots of twenty thousand or more, slips of stiff paper of the size of postcards, made up into pads of fifty or so. Have a pad always about you, and note upon one of them anything worthy of note, the subject being stated at the top and reference being made below to available books or to your own note books. If your mind is active, a day will seldom pass when you do not find a dozen items worth such recording; and at the end of twenty years, the slips having been classified and arranged and rearranged, from time to time, you will find yourself in possession of an encyclopaedia adapted to your own special wants. It is especially the small points that are thus to be noted; for the large ideas you will carry in your head. (CP 4.597)

This is another case of extended memory, of course.

26. One of Peirce’s innovations in logic is the invention of “existential graphs,” a diagrammatic version of formal logic that he considered a superior form of cognition: “A person who has learned to think in beta graphs has ideas of the utmost clearness and precision which it is practically impossible to communicate to the mind of a person who has not that advantage” (CP 4.150).

References


