**Nicolas de Condorcet: The First Philosopher of the Intelligence Explosion Hypothesis**

**1. Introduction**

Prior to Duncan Black’s rediscovery of Nicolas de Condorcet’s mathematical social choice research (Black 1958), *Sketch* *of a Historical Picture of the Progress of the Human Mind* (1795) was Condorcet’s most famous work in the Anglophone world. Thus, when works such as (Frankel 1948) and (Copleston 1960) summarized Condorcet, they focused on *Sketch*. Post-Black, there have been some wonderful pieces that applied Condorcet’s mathematical works to some of his philosophical ideas (Grofman and Feld 1988; Young 1988; List and Goodin 2001). However, these works did not focus on the central argument of *Sketch*, that human progress can be accelerated indefinitely into the future.

When we connect Condorcet’s mathematical works with his non-technical philosophical works, like *Sketch*, we find the mechanism by which he believed progress could be accelerated indefinitely into the future. In doing so, we also find that he was about 200 years ahead of his time in terms of being first to put forth a mathematical model for an intelligence explosion hypothesis.

In recent years, several philosophers (including Nick Bostrom and David Chalmers) and engineers (e.g. Ray Kurzweil, Google’s director of engineering) have extensively written about the intelligence explosion hypothesis.[[1]](#footnote-1) While there are many versions of it, the basic idea is that, in a self-sustaining way, rapidly growing knowledge and technology will radically change human economic, political, social, and biological structures.[[2]](#footnote-2) What goes unmentioned is that Nicolas de Condorcet actually created a model for an intelligence explosion in his mathematical work that explains his otherwise weak contention in his non-technical writings that knowledge can grow at an accelerating rate. Furthermore, much of Condorcet’s work in economic/political/social philosophy can be understood as an attempt to figure out how to live in a world going thru an intelligence explosion. In fact, in his last philosophical writings, he was passionately arguing that a democratic society was fundamental for a society in an intelligence explosion to sustain itself. But sadly, much of this work is unavailable to the English speaking world because many of his important mathematical and philosophical works, including *Essay on the Application of Analysis to the Probability of Majority Decisions* (1785) and *Essay on the Constitution and Functions of Provincial Assemblies* (1788) have not had complete English translations. In fact, the most complete version of his collected papers contains little of his extensive mathematical and technical writing (McLean and Hewitt 1994, 79).

In this paper, I hope to elucidate Condorcet’s model of the intelligence explosion to encourage translation and study of Condorcet’s important contributions to epistemic democratic theory and the political theory of the intelligence explosion hypothesis. Also, this paper can be considered as providing further evidence to substantiate Emma Rothschild’s assertion that Condorcet was not a cold rationalist (Rothschild 2001), Joshua Cohen’s argument that the basis of epistemic democracy is institutional (not procedural) (Cohen 1986), and Keith Michael Baker’s contention that Condorcet was not a historical determinist (Baker 1976, xxxvii). But note, I am not asserting that Condorcet’s or any other intelligence explosion hypothesis is correct; rather, I am contending Condorcet argued for a hypothesis that is only 200 years later being seriously discussed again in philosophical and engineering research.

**2. Background**

In *Sketch*, the 18th century French mathematician/philosopher/revolutionary, Condorcet, argued that human intellectual/technological/ethical knowledge growth could be faster than linear. He tried to demonstrate this by historically showing how past and then-contemporary human knowledge was increasing at an increasing rate when there existed the appropriate background conditions. *Sketch* would go on to assert that these trends could potentially occur indefinitely leading to radical changes in human social, economic, political, and biological structures (Condorcet 2004).

In response Thomas Malthus wrote *Essay on the Principle of Population, as it affects the future improvement of society with remarks on the speculations of Mr. Godwin, M.* ***Condorcet****, and other writers* (1798) (bold added for emphasis). In *Essay* (1798), Malthus accused Condorcet’s *Sketch* of assuming that past performance implies future results (i.e. past and current geometric growth do not imply future geometric growth) (Malthus 1976, 65-75).

But critics, such as Malthus, rarely if ever read Condorcet’s work on social choice, like his *Essay* (1785), where through a corollary to his jury theorem, he demonstrated conditions under which a group’s knowledge could asymptotically but quickly approach perfection (i.e. a probability of 1 of being correct). Condorcet’s jury theorem showed that if each voter had a probability greater than ½ of correctly judging whether or not some statement is true, and if each voter’s judgment of truth is statistically mutually independent of other voters’ judgments, and each voter sincerely expressed their judgment, then the majority of the voters is more probably correct in its judgment about the statement than the minority. The corollary, which I will call Condorcet’s asymptote, shows that under these conditions, as the number of voters increases, the probability of the majority being correct quickly approaches 1 (Baker 1976, 46-57).

**3. The Democratic Conditions of Condorcet’s Intelligence Explosion**

Roughly speaking, Condorcet’s jury theorem requires that one have a *large population* of voters, each of whom has sufficient levels of *knowledge*, *honesty*, *independence*, and *altruism*. Condorcet acknowledged the difficulties involved in attainment of such conditions. In fact, in *Essay* (1785), where Condorcet first expressed his jury theorem, the fourth and fifth parts of the five part book were focused on situations where the conditions were not met and also on practical applications (Baker 1976, 47-48). Condorcet’s works on philosophy, like *Sketch*, consistently attempted to advocate political and social institutions that would bring a society and humanity closer to fulfilling those necessary conditions for human knowledge to grow asymptotically towards perfection.

*3.1 Knowledge*

Condorcet contended that improving instruction and making it universally accessible was necessary in order to improve the probability of individuals being able to discern the truth (McLean and Hewitt 1994, 22-33). The purpose of instruction is twofold, to maximize the knowledge and intellect of society as a whole, and to help ensure equal rights (Baker 1976, xxviii). Condorcet explicitly states that universal instruction accelerates progress (Baker 1976, 109-110).

*3.2 Independence*

Condorcet insisted that individuals be allowed to come to their own conclusions through their own reason to help ensure statistical independence of judgments (McLean and Hewitt 1994, 37-63). For example, in an era when long distance communication and coordination were difficult, Condorcet asserted that the general population should vote by mail, rather than in a large meeting, so as to minimize the possibility of voters forming cabals (McLean and Hewitt 1994, 117, 120, 149-150). Condorcet even asserted that instruction, while potentially funded by government, should not be controlled by government, and further, instruction should be different for different students to help ensure independent thinking (Mclean and Urken 1995, xxviii). In fact, Condorcet claimed that the main reason for the speed and extent of progress in ancient Athens, more so than political liberty, was independent thinking (Lukes and Urbinati 2012, 28-29).

*3.3 Honesty*

He asserted that all persons have a right to the truth (Lukes and Urbinati 2012, 28). Furthermore, the only time deception is acceptable for Condorcet is if that deception is likely to minimize errors and deceptions in the long run (Lukes and Urbinati 2012, 78).

*3.4 Altruism*

It is clear that Condorcet believed voters needed to be altruistic, in the sense of directing their overall concerns and actions towards the general welfare of all sentient beings. However, he seems to hold the view that more enlightenment (e.g. more knowledge and capacity for reason) makes a person more altruistic (Baker 1976, 107-108). For Condorcet, all knowledge ultimately comes from sensations (Lukes and Urbinati 2012, 43). This includes moral facts and knowledge, which Condorcet asserts exist because sentient beings (capable of pain and pleasure, desire and repulsion) exist (Lukes and Urbinati 2012, 45). In other words, roughly speaking, humans have sensing capabilities, which are able to detect pain and pleasure, both in themselves and to a lesser extent in others (Lukes and Urbinati 2012, 124-125, 139-140, 198-203). Combining this information with other information, such as long term and short term trade offs, individuals can make decisions based on how the choices made probablistically affect the general welfare.

One can only wonder how post-18th century developments in mathematics and logic would have impacted Condorcet’s understanding of the relationship between knowledge and altruism. Condorcet seems to have been of the view that all truths, including moral truths, were coherently and logically intertwined (Lukes and Urbinati 2012, 122). This was a popular view in the 18th century. However, with the development of non-Euclidean geometries in the 19th century, and, for example, the 20th century discovery that the continuum hypothesis is logically independent of Zermelo-Fraenkel set theory, it is now generally accepted that mathematics is not a unified logical whole. For example, Lobachevsky geometry and Euclidean geometry are each internally consistent, but contradict each other in rejecting or accepting the parallel postulate. Perhaps if Condorcet had lived a little longer and become aware of non-Euclidean geometries, he may have come to the conclusion (perhaps à la Hume) that moral facts and natural facts are logically independent, but we will never know.

Perhaps the best argument, though weak,[[3]](#footnote-3) for why Condorcet believed knowledge growth implied moral growth, comes from Condorcet’s original solution to the collective action problem (Baker 1976, 64-66; Grofman and Feld 1988). In *Sketch*, Condorcetcontends that discovery of the truth through reason is the basis of all human progress in all realms (including, but not limited to, the fine arts, social progress, philosophy, and science). So presumably, if any human wants any sort of improvement, then pursuit of the truth leads to it. But how can a human find truth? Condorcet seems to have believed that through what we would today call crowdsourcing, humans are able to aggregate their reason and knowledge in a manner that makes them more accurate and allows them to make better decisions (Baker 1976, 283-300). This is based on the jury theorem. These better decisions are supposed to improve the lot of all humanity such that, on average, it is better for any human to live in such a society that aggregates individual opinions (e.g. a democracy) than trying to discover truth alone (e.g. some sort of state of nature). (For example, the majority of a large enough population of 51% accurate individuals under the jury theorem conditions will be more accurate on average than a 99% accurate individual, so the group will lead to greater accuracy than the individual.) But in order for this crowdsourcing of truth to work, an individual realizes that she needs to fulfill the conditions of the jury theorem insofar as possible, and she has to help others try to fulfill the conditions also. From a Condorcetian perspective, this is supposed to lead towards altruistic and honest behavior to help others and one’s self fulfill the jury theorem conditions. Thus in Rousseauian terms, the assertion seems to be that through epistemic democracy, an individual gets a share of the stag which is larger than the whole hare the individual might capture alone.

*3.5 Large Population*

Condorcet is an unusual political thinker for his times, as he held several beliefs that are now either universally or commonly held about slavery, women, sexuality, colonialism, citizen rights, and suffrage.[[4]](#footnote-4) Condorcet was an abolitionist and was very involved in the anti-slavery movement (McLean and Hewitt 1994, 341-363). Regarding women, Condorcet believed that because women are rational and sentient beings like men, they should have exactly the same rights as men (Baker 1976, 97-104). With respect to sexuality, Condorcet believed there was nothing wrong with homosexuality between consenting persons (McLean and Hewitt 1994, 56). Additionally, Condorcet argued that colonialism was wrong because it treated natives as having less rights than European colonists, though they both ought to have the same rights (Condorcet 1955, 144). Condorcet’s *Declaration of Rights* (1792) includes freedom of expression, freedom of religion, right of all citizens to vie for any public office or position, right to instruction, and the right to public aid (McLean and Hewitt 1994, 280-283). Most importantly for our purposes, Condorcet believed in universal suffrage, because in his view, an election is only legitimate if it has the consent of the people; therefore, persons whether European or non-European, woman or man, noble or peasant, have the right to vote (McLean and Hewitt 1994, 170). By seeking to decriminalize behavior and to universalize rights, universalizing the franchise increases the population of voters many fold over compared to the voting population in 18th century France. Thus, assuming other relevant conditions of the jury theorem are fulfilled, this helps ensure the probability of the majority of voters being correct approaches 1.

Additionally, Condorcet asserted that the accelerating progress of knowledge accumulation could lead to indefinitely longer lives. Condorcet was not asserting humans would become immortal, but rather for example, each year, mean life expectancy could grow by more than a year (Lukes and Urbinati 2012, 145-147). With indefinitely longer lives, as long as births outstripped deaths every year, population would grow each year.

**4. Potential Criticisms of Condorcet’s Intelligence Explosion Hypothesis**

It is clear from *Sketch* that Condorcet had put forth an intelligence explosion hypothesis, making predictions that are clearly the same or similar to those made by 21st century intelligence explosion theorists, such as (Kurzweil 2006). One, Condorcet explicitly states that intellectual and moral capacities may be indefinitely and recursively improved into the future (Lukes and Urbinati 2012, 146-147). Two, a consequence of this is the possibility of indefinitely long lives (Lukes and Urbinati 2012, 145-147).

But in *Sketch*,these assertions are based ona weak argument; basically providing a world history where progress has accelerated with the appropriate background conditions. *Sketch’s* main argument, metaphoricallyspeaking, might be understood as a non-linear regression of a scatterplot showing that knowledge has grown at an increasing rate over time. But *Sketch* does not provide a strong argument for why this faster than linear growth is occurring and being sustained. But when we look at Condorcet’s asymptote, one clearly sees a model for an intelligence explosion. By connecting Condorcet’s mathematical work (e.g. his 1785 *Essay*) with his philosophical work (e.g. his *Sketch*), we cansee that Condorcet is the earliest known thinker to model in detail an intelligence explosion hypothesis, centuries before others did so.

A few potential criticisms of this claim may be as follows. First, especially in the last chapter of *Sketch*, Condorcet makes enthusiastic assertions about the inevitability of an intelligence explosion; does this not confirm the historically deterministic interpretations of Condorcet? Second,if Condorcet’s social choice work is the model for the accelerating increases in human knowledge that Condorcet laid out in *Sketch*, then why did he not more explicitly mention it in *Sketch*? Third, if Condorcet’s intelligence explosion is dependent upon population growth to asymptotically approach perfect knowledge, is not his argument still susceptible to Malthus’ counterarguments that there are limits to population growth?

With respect to the first potential criticism, Condorcet had an empirical argument for the near certainty of a never-ending intelligence explosion; however, from a 21st century perspective, that argument seems weak. The argument goes something like this. First, states, in order to maintain themselves and stay competitive with other states, have to have some forum where sharp intellects can be developed to do physical science (Baker 1976, 21). So for examples, in classical antiquity, metallurgists and engineers were needed by states to make swords and siege engines, while in the 18th century, they were needed to make guns and canons. Even if the state wished to keep the vast majority of the society in ignorance, these forums would have to be in place to develop and preserve knowledge in general. Unfortunately, until the invention of the printing press, these forums were only able to preserve knowledge thru oral traditions and only to a limited extent, with a Photios I or Boethius, thru writing. Thus, when a state collapsed and that forum was lost, very little of that knowledge could be preserved past the collapse of that forum. In Condorcet’s view, this is why there were cycles of enlightenment (e.g. classical antiquity) and darkness (e.g. late antiquity), because after the collapse of a state, knowledge would have to be rediscovered from scratch; but the printing press changes this dynamic in two ways (Lukes and Urbinati 2012, 121). First, within each state, it allows for an increase in size of the forum, as it becomes much cheaper and easier to preserve and disseminate knowledge to more people. Second, the printing press allows enlightenment to be spread to more states, because for example, having and maintaining a printing press in one part of the world is not necessarily dependent on having printing presses in other parts of the world. Due to this non-dependence, collapse of a state (and printing presses in that collapsed state) does not destroy that hard-fought knowledge from that state, as the knowledge can be preserved in some non-collapsed state in some other part of the world. In Condorcet’s view, it would take a global catastrophe to destroy enough of the world to cause a return to darkness, but such a catastrophe was highly improbable in his view (Baker 1976, 5, 7, 15). But of course, this in the minds of 21st century readers, is a huge flaw in his argument, though it might not have been in the 18th century. In 1800, revolvers, machine guns, TNT, and nuclear weapons did not exist. The most powerful weapons were canons, which while deadly, were very unlikely to cause a global catastrophe. Condorcet may have been under the impression that more powerful weapons would not develop; though this perhaps betrays a level of naivety about progress in the development of more powerful weapons. So clearly, Condorcet has what in retrospect was a weak argument for the inevitability of a continuing self-propagating intelligence explosion, because a global catastrophe, with say for example nuclear weapons, is much more likely than Condorcet predicted. That said, Condorcet was not being mystically historically deterministic. Condorcet’s contention was based on the belief that a global catastrophe was highly unlikely, and this belief was based on empirical evidence available to him at the time. The fact that his argument might have been wrong does not make him historically deterministic.

With respect to the second potential criticism, there are a few things to note. First, Condorcet does occasionally reference the importance of voting to accelerating progress in *Sketch*, but these references tend to be fleeting (Lukes and Urbinati 2012, 59, 92, 116-117).Second, Condorcet wrote *Sketch* as a non-technical and accessible overview of his ideas that he would later expand upon in a larger work. Condorcet explicitly stated that proofs would be included in that larger work (Lukes and Urbinati 2012, 8). Thus, including math would have made *Sketch* too technical and inaccessible for the readership Condorcet wanted for *Sketch*. Third, Condorcet wrote *Sketch* under extreme duress, while he was in hiding from the French Reign of Terror. He had access to very few books and very few sheets of paper with which to write *Sketch* on. Under such conditions, never knowing with reasonable certainty if the next page he wrote would not be his last before capture, the somewhat haphazard nature of *Sketch* is understandable. Eventually the Terror caught up with him, and he was captured and sent to prison where he died under mysterious circumstances. With his death, he was never able to fill in the details with a completed larger work. Finally, Condorcet does not seem to have resolved a potential problem with his model, a problem which we today call Condorcet’s paradox, which is the intransitivity of majority preference.[[5]](#footnote-5) However, since Condorcet’s time, several scholars have shown how the problem of intransitive majority preference can be addressed (Young 1988; List and Goodin 2001; Ben-Yashar and Kraus 2002; Prasad 2012; Brams and Kilgour 2014).

With respect to the third potential criticism, Malthus criticized Condorcet’s assertion in *Sketch* that lives could be extended indefinitely. Part of Malthus’ argument was that there are finite resources that grow at an arithmetic rate, while human population would initially grow at a geometric rate; thus, resource limitations would eventually restrict this geometric population growth. It is important to note that Malthus’ criticism that there are limits to population growth was not directed at Condorcet’s asymptote, which Malthus does not seem aware of. In his *Essay on Population*, Malthus only references Condorcet’s *Sketch* and not his other works. But it is worth noting that even in *Sketch*, Condorcet acknowledges that resource limitations will slow population growth (Condorcet 2004, 74). Furthermore, in other works, Condorcet advocated for birth control (Landes 2016; Williams 2004, 168-169). But if Condorcet is advocating slow population growth, does this not slow his intelligence explosion?

It is simple to show how Condorcet’s asymptote can be slightly modified to allow for a non-growing population of voters, that still asymptotically approaches perfect knowledge. The following model is constructed to be as simple as possible while remaining close to Condorcet’s original asymptote. Additionally, this modified model allows us to show how even if voters have interdependent knowledge, this may not adversely affect information aggregation so long as votes are statistically independent. By keeping the model close to Condorcet’s, I hope to show how close Condorcet’s original asymptote was to resolving the Malthusian criticism of limits to population growth had Condorcet lived long enough to respond to Malthus. The hope is future research can modify the conditions of the result to understand how different institutions in different situations may affect the possibility and nature of an intelligence explosion. (Some of that work has already been done. The jury theorem has been generalized to allow voters to have different individual probabilities of correctness (Owen, Grofman and Feld 1989). It has also been generalized to allow for non-independent statistically correlated votes (Ladha 1992).)

The model presented here is based on Condorcet’s assertion that there is a virtuous cycle perpetuating enlightenment and instruction (Baker 1976, 5-6). This is his assertion that growth in discoveries of truths cause improvements in instruction, and that improvements in instruction accelerate the discoveries of truths. In the model, imagine a finite set of voters being presented with a series of statements. After they vote on each statement, they learn and move on to the next statement. Note, for readers not interested in the mathematics, the following section can be skipped.

**5. A Condorcetian Model for an Intelligence Explosion**

*Definitions*

Here are the definitions:

Let the set of voters be *V*, where the *n* voters are *v*1, *v*2, …, *vn* and *n* > 1.

Let the set of *m* statements be *S*: *s*1, *s*2, …, *sm*.

Any given statement *si* is in exactly one of two states: true or false.

For any given statement *si*, each voter has a probability 0 < *pi* < 1 of correctly determining the state of *si*, which is conditional on what voters learned from voting on previous statements.

The probability, that the majority of voters correctly determine the state of *si*, is *ai*.

The probability, that exactly half of voters correctly determine the state of *si*, is *bi*.

The probability, that the majority of voters incorrectly determine the state of *si*, is *ci*.

*Sincerity Axiom*

In an election on *si*, each voter votes by stating which of the two states she believes *si* to be in.

*Independence Axiom*

Define independence as follows:

Define *V*\*vj* as the set of all voters in *V* except for *vj*.

For any *vj*, how *vj* votes on any given *si* is independent of how any subset of *V*\*vj* votes on that *si*.

*Learning Axiom*

Define the learning axiom as follows:

Let 1 < *qi* < 1/*pi*

If the majority of voters correctly determine the state of *si*, then *pi*+1 = *piqi*

If exactly half of all voters correctly determine the state of *si*, then *pi*+1 = *pi*

If the majority of voters choose the wrong state of *si*, then *pi*+1 = *pi*/*qi*

Let *Q*(*V*, *S*, *p*1) specify the values of all possible *qi* given *V*, *S*, and *p*1. For brevity, we will use *Q* to refer to *Q*(*V*, *S*, *p*1).[[6]](#footnote-6)

*Discussion*

From Condorcet’s jury theorem we know that *bi*=1-*ai*-*ci*, where *ai* and *ci* are:

All *k* > *n*/2

*ai* = ∑[(*n*!/[*k*!(*n*-*k*)!])([*pi*]*k*[1-*pi*]*n*-*k*)],

All *k* < *n*/2

*ci* = ∑[(*n*!/[*k*!(*n*-*k*)!])([*pi*]*k*[1-*pi*]*n*-*k*)]

Now define *i*+1|(*pi*, *n*, *qi*) as the expected value of *pi*+1 given *pi*, *n*, and *qi*. Note that *i*+1|(*pi*, *n*, *qi*) can be expressed as: *i*+1|(*pi*, *n*, *qi*) = (*ai*)(*piqi*) + (*bi*)(*pi*) + (*ci*)(*pi*/*qi*). Furthermore, define *i*|(*p*1, *V*, *S*, *Q*) as the expected value of *pi* given *p*1, *V*, *S*, and *Q*. Using algebra, the following theorem can be proven.

*Theorem*: If (*p*1 > ½) and (for all ½ < *pi* ≤ 2-0.5, 1 < *qi* < 2*pi*) and (for all 2-0.5 ≤ *pi* < 1, 1 < *qi* < 1/*pi*), then as *m* → ∞, *m*|(*p*1, *V*, *S*, *Q*) → 1.

In other words, If (*p*1 > ½) and (for all ½ < *pi* ≤ 2-0.5, let 1 < *qi* < 2*pi*) and (for all 2-0.5 ≤ *pi* < 1, let 1 < *qi* < 1/*pi*), then as voters vote on more and more statements, the expected value of their probability of being correct asymptotically approaches absolute correctness. ■

**6. Conclusions**

The key purposes of the model are to demonstrate how close Condorcet’s original jury theorem is to modelling the perfectibility of individual intelligence, while overcoming a Malthusian critique of his asymptote and potential critiques of its statistical independence assumption. Admittedly, our new model is crude by contemporary standards. It assumes voters all have the same probability of correctly discerning the truth of statements; it assumes that voters produce their beliefs about the veracity of statements independent of one another; it also assumes that all voters express their beliefs about the statements honestly. These are highly unlikely in the empirical world. However, it emphasizes how important many institutions (advocated for by Condorcet’s social philosophy) were to help ensure voters were as honest, intelligent, and independent as possible to make sure the real world conditions were as close as possible to the theorem conditions. Future research can focus on changing the parameters, and seeing how different institutions affect the nature and possibility of an intelligence explosion.

Condorcet was 150 years ahead of his time in social choice theory and epistemic democratic theory. He was clearly about 200 years ahead of his time with respect to the intelligence explosion hypothesis. Who knows what new discoveries lurk in his mostly unexamined texts?

Such a re-examination requires important contributions from the historical and philosophical research communities. We need a more comprehensive collection of Condorcet’s writings that includes his mathematical and technical works. We need translations of his major works, most of which do not have complete English translations; even his arguably most in-depth examination of political theory, *Essay* (1788), and his foundational text on social choice, *Essay* (1785), lack complete English translations.

In contemporary times, we tend to relegate the importance of translation and re-examination of old texts for science as a relic of science during the Middle Ages, such as the recovery of Aristotle’s scientific works during the 12th and 13th centuries. What is often forgotten is how these tasks have improved knowledge and science even in relatively recent times. We know, as Benoit Mandelbrot has noted, that Abraham Robinson’s late 20th century discovery of non-standard analysis was deeply influenced by his re-examination of the works of Leibniz (Dauben 1995, xi, 51). We know Claude Shannon was introduced to Boolean logic (and thus able to establish the foundations of digital computer design), because he took a philosophy course that introduced him to the early 19th century work of George Boole (Association for Computing Machinery 2017). And of course, Duncan Black’s work on Condorcet reintroduced the jury theorem to social science.

Condorcet’s work argued that a democratic government and society is necessary for an altruistic intelligence explosion. In times where there are fears that technology will leave the overwhelming majority of people unemployed and politically disempowered in the future, a re-examination of Condorcet may lead to new insights that give us a democratic path forward.

**References**

Association for Computing Machinery. 2017. "George Boole, 1815-1864." *Association for Computing Machinery.* http://www.acm.org/about-acm/boole-bio.

Baker, Keith Michael. 1976. *Condorcet: Selected Writings.* Indianapolis: Bobbs-Merrill.

Ben-Yashar, Ruth, and Sarit Kraus. 2002. "Optimal collective dichotomous choice under quota constraints." *Economic Theory* 839-852.

Black, Duncan. 1958. *The Theory of Committees and Elections.* Cambridge.

Bostrom, Nick. 2014. *Superintelligence.* Oxford University Press.

Brams, Steven J., and D. Marc Kilgour. 2014. "When Does Approval Voting Make the “Right Choices”?" In *The Mathematics of Decisions, Elections, and Games*, edited by Karl-Dieter Crisman and Michael A. Jones, 37-53. Providence: American Mathematical Society.

Chalmers, David. 2010. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 7-65.

Cohen, Joshua. 1986. "An Epistemic Conception of Democracy." *Ethics* 26-38.

Condorcet. 2004. "Sketch for a Historical Picture of the Progress of the Human Mind: Tenth Epoch." Edited by Keith Michael Baker. *Daedalus* 65-82.

Condorcet, Marie Jean Antoine Nicolas Caritat. 1955. *Sketch for a Historical Picture of the Progress of the Human Mind.* London: Weidenfeld & Nicolson.

Copleston, Frederick. 1960. *A History of Philosophy.* Vol. Six. New York: Doubleday.

Critch, Andrew. Forthcoming. "Parametric Bounded Löb’s Theorem and Robust Cooperation of Bounded Agents." *Journal of Symbolic Logic.*

Dauben, Joseph Warren. 1995. *Abraham Robinson: The Creation of Non-Standard Analysis.* Princeton: Princeton University Press.

Frankel, Charles. 1948. *The Faith of Reason.* New York: Columbia.

Grofman, Bernard, and Scott L. Feld. 1988. "Rousseau's General Will: A Condorcetian Perspective." *American Political Science Review* 567-576.

Kurzweil, Ray. 2006. *The Singularity is Near.* New York: Viking.

Ladha, Krishna K. 1992. "The Condorcet Jury Theorem, Free Speech, and Correlated Votes ." *American Journal of Political Science* 617-634.

Landes, Joan. 2016. "The History of Feminism: Marie-Jean-Antoine-Nicolas de Caritat, Marquis de Condorcet." *Stanford Encyclopedia of Philosophy.* January 20. <https://plato.stanford.edu/archives/spr2016/entries/histfem-condorcet/>.

List, Christian, and Robert E. Goodin. 2001. "Epistemic Democracy: Generalizing the Condorcet Jury Theorem." *Journal of Political Philosophy* 277-306.

Lukes, Steven, and Nadia Urbinati, . 2012. *Condorcet: Political Writings.* Cambridge: Cambridge University Press.

Malthus, Thomas Robert. 1976. *Essay on the Principle of Population.* Edited by Philip Appleman. New York: W.W. Norton and Company.

Mclean, Iain, and Arnold B. Urken, . 1995. *Classics of Social Choice.* Ann Arbor: University of Michigan.

McLean, Iain, and Fiona Hewitt. 1994. *Condorcet: Foundations of Social Choice and Political Theory.* Aldershot: Edward Elgar Publishing Limited.

Owen, Guillermo, Bernard Grofman, and Scott L. Feld. 1989. "Proving a Distribution-Free Generalization of the Condorcet Jury Theorem." *Mathematical Social Sciences* 1-16.

Prasad, Mahendra. 2012. "Condorcet, Preference, and Judgment." *11th Meeting of the Society for Social Choice and Welfare.* New Delhi.

Rothschild, Emma. 2001. *Economic Sentiments: Adam Smith, Condorcet and the Enlightenment.* Harvard University Press.

Sandberg, Anders. 2010. "An overview of models of technological singularity." *Third Conference on Artificial General Intelligence.* http://agi-conf.org/2010/wp-content/uploads/2009/06/agi10singmodels2.pdf.

Williams, David. 2004. *Condorcet and Modernity.* Cambridge: Cambridge University Press.

Young, Peyton. 1988. "Condorcet's Theory of Voting." *American Political Science Review* 1231-1244.

Yudkowsky, Eliezer S. 2007. *Three Major Singularity Schools.* Machine Intelligence Research Institute, September 30. https://intelligence.org/2007/09/30/three-major-singularity-schools/.

1. Sometimes, the intelligence explosion hypothesis is called the singularity hypothesis (Bostrom 2014; Chalmers 2010; Kurzweil 2006; Sandberg 2010; Yudkowsky 2007). [↑](#footnote-ref-1)
2. For example, in one version, recursively self-improving AI could rapidly increase its intelligence over time in a manner that quickly outstrips human intelligence and radically changes society (Bostrom 2014). [↑](#footnote-ref-2)
3. With knowledge available today, this argument is weak for at least a few reasons. First, similar to Kant’s first formulation of the categorical imperative, it assumes what might be a false dichotomy, where the choice is between cooperating with everyone in society or cooperating with no one. There could also be choices where you cooperate sometimes and defect sometimes. Second, the Gibbard-Satterthwaite theorem and similar results roughly show that gaming of deterministic voting systems is an endemic problem that cannot be completely resolved. But given knowledge and tools available in the 18th century, the argument may have been more reasonable then. Future research in mechanism design may strengthen or weaken Condorcet’s arguments (Critch Forthcoming). [↑](#footnote-ref-3)
4. Like many people, Condorcet’s positions on issues changed over time. What I present here are the ones he held towards the end of his life. [↑](#footnote-ref-4)
5. A simple example of Condorcet’s paradox is the case with 3*n* number of voters and three alternatives (e.g. *x*, *y*, and *z*). Suppose the first *n* voters prefer *x* over *y* over *z*, the second *n* voters prefer *y* over *z* over *x*, and the third *n* voters prefer *z* over *x* over *y*. Note that majority prefers *x* over *y*, and another majority prefers *y* over *z*. If majority preference were transitive, then this would imply that the majority prefers *x* over *z*, but when we look at the preferences of the voters, a majority in fact prefers z over *x*. This was a problem for Condorcet’s asymptote that Condorcet recognized because, for example, if for any given voter and any alternatives *a*1and *a*2, a given voter prefers *a*1over *a*2 iff she believes *a*1 is more probably true than *a*2, then as *n* → ∞, Condorcet’s asymptote implies that in the case of Condorcet’s paradox with 3*n* voters that *x* is more probably true than *y* which is more probably true than *y* which is more probably true than *z* which is more probably true than *x*. [↑](#footnote-ref-5)
6. Informally speaking, note that the path of the society of voters through the statements can be visualized with a ternary tree, where each non-leaf node sprouts exactly three children nodes: the majority is correct, exactly half are correct, and the majority is incorrect on the given statement. So if *S* has *m* statements, *Q* specifies 30+31+…+3*m*-1 possible *qi* values. This is because, though *Q* generates 3*i*-1 possible *qi* values for any given *ai*, which one of those 3*i*-1 possible *qi* values is the one that is actualized is dependent on the path the society of voters takes from the root to the depth of *ai*. (Each node at depth *i*-1 is at the depth of *ai*). Thus, we could prove this theorem with structural induction. [↑](#footnote-ref-6)