

**New Media, Speculative Technology, and Changing Conceptions of Citizenship**

**(formerly: Yelping as a Citizenship Act)**

**Matthew J. Moore**

**WPSA 2016 – San Diego, CA**

**Conference Draft—Please do not quote without permission.**

Matthew J. Moore  
Associate Professor  
Dept. of Political Science  
Cal Poly State University  
San Luis Obispo, CA 93407

mmoore02@calpoly.edu

To paraphrase Wittgenstein: I had hoped to write a good conference paper, but the time for that has now passed. Instead, I invite you to workshop some very preliminary ideas with me. The first set of ideas has to do with the possibilities that new/social media offer for rethinking citizenship and improving democratic engagement. The second set has to do with other changes and challenges to citizenship and social cooperation that new and speculative technologies pose.

### Part I: New/Social Media and Citizenship: Is Yelping a Citizenship Act?

From one perspective, much new media looks like a culture of public narcissism, an endless competition of retouched photos, clever ripostes, and conspicuous consumption. But from another perspective, new media (and some old media whose relevance has risen on the new media tide) are surprisingly neighborly and civic minded, at times approaching and perhaps even crossing the threshold of being acts of citizenship.

As a first example, take the social-media website Yelp, on which semi-anonymous users post reviews of restaurants, bars, shops, services, and even government agencies. Presumably some reviews are simply fraudulent, and no doubt others are written mainly as expressions of ego or vanity, simply to hear oneself talk. But the vast majority appear to be exactly what they claim: honest if opinionated reviews posted by customers with no personal stake in the businesses. It's worth taking a step back to ask why so many millions of people write these reviews. After all, they aren't paid to do so; the reviews are semi-anonymous, such that the authors are unlikely to get any personal benefit from writing them; and given the traffic on Yelp, the readers are likely to be perfect strangers, whom the reviewers have no special reason to help.

The answer seems to be some combination of enlightened self-interest and civic-mindedness: people write reviews because they themselves want access to such reviews and realize that reciprocity is necessary to sustain a reviewing culture, and /or because they are just willing to help the general public.

That combination of motives—enlightened self-interest and civic-mindedness—is also what (ideally) motivates citizenship. When someone obeys an inconvenient law, pays their taxes, shows up for jury duty, votes or otherwise acts to affect policymaking and implementation, reports a public nuisance or hazard, or in some other way fulfills their duties as a citizen, they do so, in the best case, because: (a) they judge that the benefits of peaceful social cooperation are worth the costs; (b) they recognize that the cost of securing the cooperation and obedience of others is their own cooperation and obedience; (c) they are motivated by goodwill to benefit the generic public other, by making some shared resource better than it would otherwise be. (Of course some citizens may do their duty primarily due to fear of punishment, and of course some citizenship acts (like attending a public meeting or hearing) may primarily be aimed at securing some private benefit (a favorable law or regulation, for example).)

Yelping is most obviously similar to (or an example of) a citizenship act when the review is aimed at evaluating government itself. Thus Yelpers review DMV offices, passport offices, Post Office branches, libraries, courthouses, and other public institutions, and many reviews contain helpful tips (can you eat in the waiting room? are there diaper changing tables in the restrooms? do you need an appointment? is there someone on staff who speaks Hmong?). Yelp could also be a valuable conduit for publicizing government-generated information, such as the results of health department inspections of restaurants.

The most striking difference between the civic-mindedness of Yelping and the civic-mindedness of more obvious citizenship acts like paying taxes, voting, jury duty, and obeying inconvenient laws is that millions of people Yelp voluntarily,<sup>1</sup> while it is notoriously difficult to get many citizens to perform their explicit duties willingly. Could we learn something from Yelp about ways to rethink citizenship, to broaden our understanding of it, to make engaging in it more appealing?

### *The Green Book*

One likely objection to looking at Yelp and other allegedly pro-civic new media for ideas about how to rethink or encourage citizenship is that while there may be some similarities or analogies, the dissimilarities are so significant as to make the comparison meaningless. Thus, citizenship often requires people to sacrifice things (time, money, even physical safety or one's life, in the case of conscription) that they value greatly. Yelping requires only a small amount of time. From another perspective, arguing that Yelping is a citizenship act is to neoliberalize citizenship by turning it into an individual act, done in private, aimed at improving the experience of consumption. Such a comparison would be to trivialize and depoliticize citizenship.

To try to make the case for seeing some new media behavior as citizenship acts, I want to invoke an old-media phenomenon that is similar to Yelp and yet also clearly related to citizenship. Beginning in 1936 and continuing until 1964, Victor Hugo Green, an African-American postal carrier in New York City, along with his wife Alma Green, published *The Negro Motorist Green Book* (in later editions called *The Negro Travelers' Green Book*, and then

just *The Green Book*), a travel guide that detailed which businesses would serve blacks during the Jim Crow era.<sup>2</sup> The first edition was 16 pages; by 1962 the *Green Book* was almost 100 pages long, was available at Esso gas stations across America, and was selling two million copies a year.<sup>3</sup>

Green asked fellow postal carriers for information, but he also relied on reports from travelers. The 1937 edition included this appeal:

We are appealing to the Motorist and Business places for their whole-hearted co-operation to help us in our endeavor, by contributing ideas, suggestions, Travel information, and articles of interest....In the event that you are dissatisfied with the service rendered by an advertiser, we would appreciate you writing us the complaint...and we will immediately investigate....Let's all get together and make Motoring better.<sup>4</sup>

Obviously, the information in the *Green Book* was more important than virtually any of the information on Yelp, since it protected black motorists' physical safety and ability to travel, and did not merely tell them where to get the best taco in Tulsa (Baja Jack's Burrito Shack, according to Yelp). But structurally the *Green Book* was a pre-internet Yelp, allowing would-be consumers to share helpful information about businesses. It differed from other pre-internet media in that it was not merely an encyclopedic listing of every relevant business (like a phone book or AAA travel guide), nor was it simply a venue for paid advertisements from any business with ready money (like a Chamber of Commerce guide). Rather, it was a collaborative project aimed at allowing a particular group of people to share information that they wanted but had no

other way of obtaining, and allowed anyone to contribute to or correct that information, merely because they valued the project and wanted to help.

And it seems difficult to deny that the publication of the *Green Book*, and the efforts to update it made by thousands of travelers with good or bad stories to tell, were civic-minded acts by people who combined enlightened self-interest with a desire to help others who faced similar problems. In other words, they were citizenship acts. That the Green Book contributed to the struggle against racial discrimination certainly makes it a more noble activity than Yelp, but it doesn't make it a fundamentally different activity. In both cases, the combination of enlightened self-interest and civic-mindedness turns what initially appears to be a private, consumption-oriented action into an act of citizenship, however modest.

### Three Models of Citizenship

Since I'm making claims about what counts as an act of citizenship, it might help to take a step back and talk about what I mean by the term. Scholarship typically identifies two broad, alternative conceptions of citizenship: republican and liberal. On this view, the republican conception of citizenship rests on citizens willingly subordinating their individual interests to the collective interest of the society. For example, Rousseau's sovereign can only be formed by each citizen giving up all of his or her rights to the collective, and it can only function properly if it acts based on the General Will (the policy that is maximally in the collective interest) and not on the private wills of individuals, either singly or aggregated as the Will of All. Republican citizenship is threatened by the loss of civic virtue, such that the citizens seek their individual or factional interests, and in the process possibly undermine the basis of social cooperation.

In contrast, liberal citizenship rests on citizens pursuing their individual interests in a way that is socially harmonious, either because their self-interest is enlightened enough for them to perceive that continued cooperation is more valuable than a short-term gain that undermines the system (Kant) or because the institutions of social cooperation are constructed in such a way that the various self-interests cancel out (Madison) or harmonize (Adam Smith). Liberal citizenship is threatened by a failure of self-interest to be enlightened (perhaps due to short-sightedness), the emergence of factions that can act collectively and thus avoid the balancing-out of individual wills, and market failures (asymmetries of information or bargaining power, externalities, etc.), which distort the costs and benefits of collective action.

As I suggested above, in American popular culture the prevailing conception of citizenship seems to contain both liberal and republican elements. Thus, although we expect that people will generally vote according to their preferences rather than according to what they imagine the general will to be or what is in the collective best interest of the society, we also expect that citizens will be willing to put some limits on their self-interest for the sake of preserving not merely a system of social cooperation, but a fair system. Thus as a society we have approved and currently practice various forms of economic redistribution, we pay taxes to support public schools even when we do not have children attending them, and otherwise spend our time, money, and attention to help the generalized public other. And we expect that people will do this in the spirit of generalized reciprocity (for example, paying in based on capacity but drawing out based on need) and not specific reciprocity (expecting to get back pretty much the same as we pay in).

As I suggested above, Yelping (and other new media activities, like posting reviews on Amazon, open-source software, and Wikipedia) appear to reflect exactly the combination of

enlightened self-interest and civic-mindedness that we see in the hybrid conception of citizenship that prevails in American popular culture. That suggests two lines of further inquiry: (1) should we reconceptualize citizenship to include acts like Yelping? (2) could we draw lessons from the success of activities like Yelping to increase and improve democratic engagement? In the next two sections, I consider two recent books that attempt to do both of these things.

### *Citizenville*

California Lieutenant Governor, and former San Francisco mayor, Gavin Newsom is an evangelist for the idea that some new media is similar to citizenship activity, and that government could dramatically improve by learning some lessons from the digital world. His 2013 book *Citizenville* rests on a contrast between two metaphors for government. The pre-digital metaphor for government was a vending machine—you put in some money and out comes a product or service. If the machine just eats your money, you can shake the machine. Getting it stocked with new options is a lengthy, tedious process that has a very high likelihood of failure. In contrast, the reigning metaphor for new, digital media is the cloud—real-time access to information and services, from anywhere at any time, and with a high degree of both flexibility and two-way communication. See a mistake on Wikipedia? Correct it yourself in a few minutes. Get bad service at a business? Yelp about it and see the response in real time, in the form of replies from the business, other reviews, and even news coverage. Find a bug in open-source software? Report it so that others can verify and fix it, or fix it yourself and make your improvement available to the world. Newsom argues that government needs to become more like the cloud and less like the vending machine.



What would that look like in practice? Newsom identifies five key changes: “First, government has to be absolutely transparent—every agency, across the board, with exceptions only to protect public safety or personal privacy. We must open up our vast stores of data...and make sure they are standardized and easy to use. Second, we must encourage people to use that data to create useful apps, devices, tools—anything they want....Third, we must learn to engage people on their own terms....Fourth, we need to allow people to bypass government. We must encourage them to take matters into their own hands....Fifth, we must inject a more innovative, entrepreneurial mind-set into government” (Newsom xxi-xxii ).

Achieving those goals requires a re-imagining of government’s role in society: “We have to think of government as a convener. A concierge, if you will—a customer-service organization” (Newsom 81). He proposes a variety of specific policy changes to achieve that re-conceptualization, such as creating a system of flexible, reliable on-line voting (Newsom 162), using contests and innovation-prize competitions, creating public-service review applications (Rate My DMV), and creating government app stores for accessing and using government data (Newsom 232). He also cites a number of interesting and even inspiring examples, including:

Code for America, a non-profit that pairs volunteer coders with public agencies that need help overhauling existing systems or creating innovative new ones. (Newsom thinks projects like this are especially important because government will probably never be able to pay top tech talent enough to lure it away from the private sector on a permanent basis.)

Open 311, an open-source platform with defined data standards to allow government agencies to make data publicly available in universal formats, so the data can be easily used, compared, and so on.

Manor Labs, a game-like website created by Manor, Texas, on which citizens competed for virtual rewards by reporting real problems in the city, like potholes or overgrown trees, as well as for proposing innovations (innovations actually put into practice paid bonuses).

Project Madison, which basically opens up the process of drafting legislation to the world, Wikipedia-style. This is a particular case of the broader open-source government movement, which seeks to translate the success of the open-source software model to the legislative process.

In addition to the examples Newsom identifies, there are numerous similar efforts:

Myfairelection.com -- A website set up by Harvard political scientist Archon Fung to be a kind of Yelp for polling places in 2012, aggregating reviews and reports of either good or bad practices.

Votesquared.org – A social platform for ranking politicians.

Challenge.gov – A clearinghouse website for government-sponsored competitions and innovation prizes.

Newsom ends the book on an inspiring note: “It’s not that technology strengthens government. Instead, technology changes the very nature of government---from a top-down entity to a bottom-up one. From one-way hierarchy to two-way democracy. From the vending machine to the cloud...The democratizing influence of the cloud leads to a stronger, more stable commonwealth” (Newsom 233).

### *Making Democracy Fun*

Newsom touches repeatedly on the idea of using games and game-like design to increase public engagement with government. In *Making Democracy Fun*, Josh Lerner examines that idea in much greater depth. Lerner provides two detailed case studies of the use of games to increase democratic participation and stakeholder buy-in—the process of rectifying the urban design and property titles in a shanty-town in Rosario, Argentina; and an annual competition for tens of millions of dollars for repairs and improvements in public housing developments in Toronto. He also provides a nuanced and detailed argument about how game elements can and cannot improve democracy.

Building on a framework suggested by Colin Hay, Lerner is interested in investigating not the demand-side of political engagement (why are people decreasingly willing to engage?), but rather the supply side (are the processes of democratic engagement worthy of citizens’ time and trust?). His central claim is this: “For most people, democratic participation is relatively

unappealing. It is boring, painful, and pointless” (Lerner 13). Games, in contrast, are engaging, pleasurable, and purposeful. Games are “systems in which players engage in an *artificial conflict*, defined by *rules*, that results in *measurable outcomes*” (Lerner 16).

Lerner proposes that there are five kinds of games that could be democracy-enhancing:

Animation – “*Animation games* transformed passive observers into active participants, usually at the beginning of meetings and workshops” (Lerner 190).

Team-Building – “*Team-building games* integrated active individuals into collaborative teams....Rosario Hábítat’s *puzzle challenge* taught *villa* residents that they had to communicate clearly within *and* across groups to find the missing puzzle pieces—and to fix neighborhood problems” (Lerner 190).

Capacity-Building – “*Capacity-building games* armed people with the skills and knowledge necessary to work together” (Lerner 190).

Analysis – “*Analysis games* offered accessible ways to understand complex systems, by breaking these systems into manageable chunks....In Toronto, we used the *Matching Puzzle* to break down another daunting topic—research indicators. Only by mixing and matching indicators with real-world examples did the tenants grasp how our indicators played out in practice” (Lerner 191).

Decision-Making – “*Decision-making* games enabled players to harness their new knowledge, skills, and understandings to reach concrete decisions” (Lerner 190).

In addition to these specific types of games, Lerner advocates incorporating game dynamics into democratic engagement. For example, all games include some kind of *conflict*, and game designers are careful to frame and manage that conflict to make for the best game. Thus the conflict might be individual vs. individual, individual vs. system, individual vs. group, group vs. group, or group vs. system. Often it’s possible to reframe conflict from one type to another, thereby turning an adversary into a collaborative process. Similarly, game designers are very deliberate about how *rules* are made, communicated, and enforced. *Outcomes* is a third game dynamic—no one would want to play a game in which the outcomes were unclear or constantly changed, and we should expect the same reactions in democratic engagement. Further, game designers are careful to provide *feedback* so that players can understand whether they are making progress or not, and how much further they have to go. Games are carefully crafted to maintain the *engagement* of players, through strategic use of rewards, temporary setbacks, opportunities to make choices, audio-visual stimulation, and the use of narrative.

Although Lerner acknowledges that games could be bad for democracy by possibly leading to violence, trivializing issues, manipulating participants, leading to unfair outcomes, or simply not being fun (Lerner 19), he ends on an enthusiastic note: “Despite its many merits, democracy is not the most enticing pastime for most people....When democracy is unappealing, it attracts a rather unrepresentative lot—political junkies, passionate extremists, and other usual suspects....By making democracy fun, we can make it work. Democratic participation can and must be more than a civic chore” (Lerner 207).

## Functions of Government

On one level, Newsom and Lerner's enthusiasm for the possibilities of leveraging new media and games to improve democratic engagement is both understandable and even persuasive. Surely many people have looked at the astonishing success of Wikipedia or open-source software and thought: "If we can get millions of people to dedicate countless hours to these things, why can't we do them to do the same for government?" Conversely, presumably many have also thought: "If we can get libraries everywhere to systematize and share their data through something like WorldCat, why can't we get government agencies to do the same thing?"

Above I briefly suggested that one difference between government and new-media entities like Yelp involves differences of scale. Yelp involves small amounts of time, whereas jury duty, attending city council meetings, or being conscripted into the military involve large amounts of time. Similarly, since many new-media products generate income through advertising, they are often free or very inexpensive to the end user, whereas government requires substantial contributions of money from a significant portion of the society. (Interestingly, it turns out that Mitt Romney was not far off when he asserted that nearly half of Americans pay no income tax, though of course they pay other taxes like sales tax, property tax, excise taxes, and so on.) Surely Newsom and Lerner are right that there is substantial room for improvement in the average citizen's experience of non-optional citizenship duties (making election day a holiday, online voting, simplified tax forms, etc.). But citizenship will always require more of citizens' valued resources than Yelping will.

Here I want to suggest that there is another disanalogy between government and new media: government necessarily engages in some activities that are intrinsically unpleasant, and that no amount of graphics or peppy music can make desirable. As a preliminary list, it seems to me that government engages in several related but distinct kinds of functions:

Enabling: Creating a process for obtaining something optional but potentially desirable, like a legal status (marriage, child custody, incorporation) or the conduct of private business (contracts).

Extractive: Obtaining money and person-hours to accomplish the work of government / the public.

Deliberative: Creating and maintaining a process by which the public can deliberate on matters of common concern and come to decisions.

Supervisory: Watching over certain activities on behalf of the public—business regulation, child welfare policies, licensing, auditing government's own actions.

Punitive: Imposing penalties for violations of rules / conflicts with the public will.

Coordinating: Creating institutions and processes for taking collective action.

Based on that preliminary division, it seems likely that some functions of government are going to be more amenable to increased participation through cloud-ification and game-ification than others. Thus, the deliberative function of government seems open to change. Online voting, transparent availability of election data, game-inspired forms of local deliberation and decision-making all seem like possible (and perhaps even desirable) changes. The enabling function of government is another obvious site for reform, since in that area government is offering things that citizens want and can't get elsewhere. Thus government might improve its customer service by creating ratings apps for the DMV (or encouraging citizens to rate agencies through existing services like Yelp), making information more easily available and usable by the public or third-parties (health inspections of restaurants, real-time locations of transit vehicles, simplified legal forms or processes). Similarly, both the coordinating and the supervisory functions also seem to be amenable to change, for example by making data available, by making coordinating functions participatory and responsive, and so on.

In contrast, government's extractive and punitive functions seem much less susceptible to the kinds of improvements Newsom and Lerner envision. That's not to say that there is no room for change or improvement. For example, the use of restorative justice techniques (face to face meetings of offender and victim, involvement of the offender's family and community in punishment and rehabilitation, analysis of the social sources of crime) in some limited areas (juvenile justice, petty crime) moves in the direction of greater democratic engagement and government responsiveness. Further, in both areas there is data to be made accessible and usable, transparency to be achieved, and so on. And certainly democracy-enhancing improvements anywhere in government would probably make citizens more willing to contribute the time, money, and perhaps risk of harm that citizenship requires. But it's hard to imagine that any



change will make citizens eager to file their taxes, pay fines, or appear in court. More generally, it seems obvious that it will be much easier to make government more engaging and responsive when it is helping people do something they want to do anyway, and much harder when government is requiring people to do something that they would prefer not to do. Reform is easier where citizenship coincides with self-interest, harder where citizenship requires enlightenment and civic virtue. This suggests that there may be limits to how much we can learn about citizenship from Yelp.

## Part II: Speculative Technologies and Citizenship

To avoid you thinking that I've completely gone off the deep end, let's start this discussion with a speculative technology that you've heard of and that is reasonably close to being a reality: self-driving cars. Think with me for a few moments about how self-driving cars will impact society. First when would it be ethically responsible to adopt self-driving cars, given the stubborn bugs in their systems? I want to suggest that the answer is fairly obvious: when replacing human drivers with self-driving cars will result in fewer driving-related injuries and fatalities than we experience with human drivers, even if we know for a fact that self-driving cars will kill or seriously injure x number of people per year because we haven't perfected them yet. I'm not a thorough-going consequentialist by any means, but since every realistic course of action involves injuries and deaths, minimizing those bad outcomes seems like the obvious criterion for making a choice. From what I can tell, it seems like we are already very close to this point if we haven't already reached it.

When those inevitable injuries and deaths occur, who will be liable? You don't have to be a defense attorney to realize that no company is going to sell self-driving cars if it will be liable for damages caused by the ordinary operations of the cars. Surely the social price of self-driving cars will be immunity from liability for their manufacturers (assuming they meet some standard of care in design and production). But if I'm riding in a self-driving car that causes harm to someone, it's hard to see how I could be personally liable. If I am neither driving the car nor programming its systems, I'm in no position to prevent the harm; indeed, I'm effectively in the same position as a passenger on a train or bus. Yet of course the victim of the accident needs to be compensated. My guess is that the result will be universal no-fault insurance. Everyone has insurance, and each person's policy covers their injuries and damages, without attempting to establish fault or recover the costs from someone else. That change, in turn, would result in a significant reduction in the demand for the services of lawyers. Human-driven vehicles might still exist, but presumably they would be very expensive to insure, since humans are so much more dangerous as drivers than computers, and thus might be luxury items affordable only by the wealthy.

Though, now that we're on the topic of car ownership, why would people own cars at all? The average car spends roughly 90% of its working life parked and doing nothing but losing resale value and taking up valuable real estate. If your car could drop you off and then come back and pick you up, there would be no reason to have parking located in areas popular with humans. Why not put the parking lots way out on the outskirts of town? But for that matter, why have parking lots at all? If your car will come when you call, why not let it work for Uber or Lyft when you don't need it? It's time our cars earned their keep! But, once again, if your car is spending 90% of its time driving other people around, why should you shoulder the burdens of

owning it—the insurance, repairs and maintenance, fuel, garaging, and so on? Wouldn't it make more sense to join some kind of cooperative (or patronize a business) that allowed you to be driven in whatever car you wanted whenever you wanted, and charged per trip?

Think with me for another few moments about how self-driving cars would affect other areas of life. Presumably designated drivers would be a thing of the past, as would drunken driving. That can only lead to a significant increase in how much people drink. Or what about driving children to and from school and enrichment activities? Assuming that self-driving cars could be programmed with various safety features—only opening their doors at a programmed destination, or in response to a preset code, for example—why not let the car drive the kids to soccer? And why not let the car drive Grandma to lunch with her friends, or to the grocery store or her doctor's appointment? If you could work or sleep while the car drives you to and from work, how long a commute would you be willing to endure? Would self-driving cars lead to increased sprawl, especially considering that they will be able to drive safely at higher highway speeds than human drivers, and in much closer proximity to other cars, thereby possibly relieving traffic congestion and dramatically speeding up commutes in some areas? Would the reduced opportunity costs of driving lead to much more driving, and correspondingly greater outputs of pollutants and greenhouse gasses?

My point here is that we can imagine / foresee that the almost-here technology of self-driving cars will almost certainly have profound effects on the legal, social, economic, and ultimately political structures of our society. What about more speculative, less imminent (and perhaps less feasible) technologies that are currently being discussed? Here I want to touch on three: robots, artificial intelligence, and whole brain emulation.

## *Robots*

Of the three, robots are the easiest. By robots I mean machines that are capable of replacing human labor, but are not artificially intelligent—that is, they are capable of following directions, even vague or complex directions, but they are not capable of independent learning (other than in a very limited sense), choice, self-reproduction, self-reflection, and so on. In a word, they aren't sentient. As with self-driving cars, the real issue with robots is the question of how their existence will change human habits and patterns. For example, if a large portion of the labor that is today done by people can be done by robots in the future, what will the displaced workers do to fill their days and earn their livings? Will we need to pay a universal basic income to humans whose labor is no longer needed, as some in Silicon Valley have recently suggested? (Manjoo) Are humans well adapted to endless leisure and having no meaningful work, or will we end up like the people in *Wall-e*? Will social power inevitably shift to a technocratic elite made up of the people whose labor and talents are still necessary to creating and maintaining the robots, as Theodore Kaczynski (aka The Unabomber) argued (Kaczynski Sec. 174)? Will divorcing our ability to consume from our ability to work lead to runaway consumption, or an inability to grasp the environmental or social costs of our consumption?

Another disturbing set of questions involves our relationships to the robots, and through them to labor, and ultimately to each other. Basically, the robots would be slaves, but because they aren't sentient, we don't have to feel guilty about enslaving them. On its face, that's not deeply troubling—we could equally accurately describe the robots as tools or machines. But it seems likely that pushing all boring, tedious, dangerous or difficult work onto robots would reproduce the mindset that Tocqueville identified in the pre-Civil War slave states: that labor was

slavish, and leisure (sometimes indistinguishable from torpor) was for free people. Would children refuse to clean up their toys or help with the dishes, on the grounds that “that’s ’bot work”?

More disturbing is the changes that are likely to arise in human relations with each other. It seems obvious that if we develop the ability to create robots that can replace human labor, then any human activity that someone is willing to pay for will eventually be done (at least in part) by robots, and that means that both sex and romance/companionship will eventually be available from robots. Again, we don’t need to imagine artificially intelligent robots for these roles—even today we have the ability to create computer programs that can carry on natural-language conversations convincingly enough that, at least in some settings, people don’t know they’re not talking to a person. And, conversely, we know from the universality of prostitution and escort services that people are willing to pay for mechanical sex and the illusion of companionship. I’ll let you use your own imagination to fill in the lurid details, but the basic point is that people will develop sexual and romantic/companionate relationships with human-like robots that will obey their every command, and that do not demand recognition of their subjectivity or agency (since they have none). That seems likely to introduce profound tensions into human-human relationships. To make the point in broad brush strokes: if the stereotypical heterosexual man had a human-like robot that would do all the domestic work, satisfy all of his sexual desires, and never ask for anything in return, would he ever pursue a relationship with an actual woman? And if he did, would his expectations be so unrealistic as to undermine any fledgling connection? What if the robot were capable of carrying on light conversations about current events, celebrities, and the man’s interests and hobbies? What if the robot were good (but not too good)

at video games and poker? And of course the same general line of reasoning applies to women and their Stepford husbands, gay men and women, and so on.

Despite the deeply disturbing possibilities such robots raise, they're the easy case in terms of the three speculative technologies I'm interested in, because the robots aren't sentient, and thus we don't have to work out new moral and political relationships with them. Rather, they're merely an extremely disruptive technology.

### *Artificial Intelligence*

By artificial intelligence (AI) I mean computer systems that achieve sentience. It turns out that not only are we quite far away from such capabilities at present, but we are not even entirely sure how to define sentience. Without getting into the weeds of that debate now, by sentience I mean roughly human-like capacities for learning, deliberation, choice, self-reflection, growth and reproduction. The basic political question regarding AI is whether artificially intelligent entities should be enfranchised. This is where the parallel between robot labor and slaves becomes morally urgent. While it is morally acceptable to treat as things non-sentient machines that do a great deal of labor, it is at the very least questionable whether it would be acceptable to treat sentient machines that way. Rather, creating sentient entities to do our work, but refusing to give them social or political power or equality, despite what I think we can assume would be their requests for such rights, would be to literally recreate slavery.

But the prospect of giving social and political rights to AI entities (AIEs) raises a series of extremely difficult problems, all of them revolving around the ways in which such entities would not be similar to human beings. (I put aside for present purposes the extremely difficult

question of how we would know whether a computer program has achieved sentience—the famous Turing test is obviously inadequate, as John Searle’s Chinese Room counterexample demonstrates.) First, it’s not obvious that AIEs would experience emotions, or if they did, whether their emotions would be relevantly similar to human emotions. There are three reasons to suspect that AIEs would not experience human emotions: (a) it’s not obvious that human programmers seeking to create AI would include a capacity for emotions, since they seem likely to distract AIEs from cognitive tasks, just as emotions do for humans; (b) it’s not obvious that AIEs would, in the process of their own growth and self-development, choose or be able to develop a capacity for experiencing emotions, since doing so would introduce a quality of randomness into an otherwise predictable and reliable system; (c) finally, it’s not obvious that AIEs that were given or developed emotions would choose to keep them, for the same reasons. But emotions are central to what it means to be human, and are therefore central to what it means to humans to be a political actor. Beings that do not experience fear, dread, hope, excitement and so on may genuinely have a moral interest in being able to achieve peaceful social cooperation with each other and with human beings, but it seems likely that their interests, preferences, and reactions to events would often be dramatically different than those of humans. It’s not even clear that AIEs and humans would perceive the same circumstances as giving rise to the same issues to be decided. For example, would inorganic AIEs be concerned about the conservation of the ecosphere, or about the extinction of particular species?

A second dissimilarity is that AIEs would effectively be immortal, since everything that defines and is distinctive about them could be identified, copied, reproduced, and transferred. Perhaps it bears pointing out that we haven’t even the faintest idea how to do this with human beings (but see the next section for more on that issue). Just as with emotions, so mortality is

central to what it means to be human. The facts that we have limited time to spend, that we are susceptible to illness and death at any moment, and that if we are lucky we will live long enough to watch our capabilities fade into senescence, define human existence. What gives our lives narrative coherence and meaning is that they represent an endless series of choices under conditions of scarcity—we are constantly slamming shut doors of possibility because that is the only way to walk through any one of them. In contrast, an immortal being never fully gives up on a possibility, but merely postpones exploring it until later. Such entities and human beings would face dramatically different choices, even when ostensibly deciding the same issues. For example, a public policy whose costs are felt now but whose benefits would accrue hundreds of years in the future would require humans to exercise intergenerational altruism, which we generally find quite difficult, but would require AIEs only to recognize that they will reap the benefits later that they have paid for today. Similarly, the decision to wage war poses an existential threat to human beings, but not necessarily to AIEs (though presumably even they would be subject to various attacks and dangers of permanent destruction, giving the metaphor of a computer virus a new poignancy).

Third, AIEs are capable of nearly infinite reproduction at a very low marginal cost (at least at the moment; energy consumption and raw material for building computers would ultimately be the limiting factors). If we adopted the policy of one-sentient-being-one-vote, AIEs could easily outnumber humans and win every election or decision. Further, if AIEs reproduced by cloning (that is, making exact replicas of themselves), AIEs might tend to have similar or even identical interests and preferences (at least for a while, until time and growth allowed greater diversification), making them a monolithic faction. Or, given that AIEs would probably be able to learn at a nearly infinite rate, it might be that each AIE's political preferences changed



wildly from election to election, making forecasting and coalition building essentially impossible.

Fourth, it seems likely that AIEs would engage in different moral reasoning than do humans. Human moral reasoning seems to be driven partially by moral intuitions and partially by logic (with healthy doses of emotion, disguised self-interest, and confusion thrown in). The famous Trolley Problem suggests that most people have conflicting moral intuitions, with deontological intuitions predominating in some scenarios, consequentialist intuitions predominating in others, and a substantial gray area in which neither clearly predominates and we have no clear idea how to justify our decisions against their rival claims. How might AIEs reason morally? It seems to me that there are only a few possibilities: (a) they will reason just like we do, and thus be as confused as we are; (b) they will develop a logically compelling argument for either deontology or consequentialism, which humans will recognize as being logically sound, but which most of us cannot consistently follow because of our conflicting intuitions; (c) they will simply choose either deontology or consequentialism and pursue it consistently, which most of us will not be able to do; (d) they will adopt some new moral system that we can neither understand nor follow; (e) they will be unable to understand moral reasoning, viewing it as a human eccentricity. Given that range of possibilities, unless they reason like we do (presumably without the feelings of inadequacy, tragic compromises, dirty hands, and so on that we experience), we and they will frequently come to very different moral conclusions and consequent policy preferences, giving rise to the familiar problems of moral pluralism.

There is one other issue that poses a problem and that I want to mention, though properly it is part of the questions of whether AI is even possible in principle and how we would know whether a computer had achieved AI. That is the question of free will. Remember that ultimately

an AIE is a computer program—it is a set of instructions and data, stored on physical media. There can be no metaphysical speculation about where an AIE's soul is. Off the cuff, it seems to me that logically there are three possibilities about how AIEs will reason: (a) they will reason deterministically, such that every outcome is determined by its antecedents; even in the case of probabilistic decisions or situations, the AIE would have explicitly stated thresholds for choosing among possible outcomes; (b) they or their programmers will intentionally introduce randomness into their reasoning, presumably because experience has shown that a certain degree of randomness leads to an overall better outcome (perhaps randomness is a proxy for creativity or inspiration); (c) randomness will be introduced into the system unintentionally, by things like damage to computers, interference with data transmission, errors in the code, quantum fluctuations in components, and so on. The question is whether any of these scenarios counts as free will. It's hard to see how (a) could be called free will, since by hypothesis every choice is determined by its antecedents, and in principle it would be possible to predict every future choice if we had access to the initial code and could foresee the empirical phenomena the AIE would encounter. Possibilities (b) and (c) seem problematic as well, for the different reason that randomness is not the same thing as will or intention. Imagine that there existed a person whose behavior was entirely deterministic and predictable, but whose behavior could be changed if you pressed a button that introduced an element of randomness into his decision-making process. His behavior would no longer be predictable, but I don't think we would say that his future behavior was due to his own will, intention, or choice. Indeed, once we knew what the random element was, we could again predict his behavior perfectly until the next button push. That's not what we (currently) mean by free will (not that we have a terribly clear idea of what we do mean, but this isn't it). If my reasoning here is correct, this means one of two things: either, (a), human beings

do genuinely experience/possess free will but no AIE is even in principle capable of achieving it, such that no AIE is capable of real sentience (despite what might be compelling evidence of apparent sentience) and thus we will never face the moral and political questions of whether to enfranchise AIEs; or, (b), humans also do not possess free will, and our illusion of free will arises from some form of randomness or from our inability to understand or analyze the complex deterministic factors that lead to our behavior, such that we may indeed need to treat AIEs as moral and political equals, while also nursing our own bruised self-conception.

In science-fiction, the problem of AIEs is usually cast in terms of whether we will be able to develop human-AIE relations that the AIEs will accept without revolting, killing or enslaving us all, or otherwise taking over and becoming the new dominant species. Here, I'm interested in the converse question: can we develop human-AIE relations that humans can rightly judge as being morally acceptable, without creating a situation in which either humans are subordinated to AIEs or we discover that co-existence is impossible? The dissimilarities discussed above suggest that there are profound difficulties in the way of achieving just human-AIE social cooperation.

### *Whole Brain Emulation (aka Mind Uploading)*

If you thought the AI section was crazy, here you'll think I've completely lost it. A number of futurists and technology experts believe that it will eventually be possible to create computer programs that perfectly emulate the human brain, such that it would in principle be possible to copy or upload an individual's consciousness to a computer. Again, this issue is hotly debated, and for the moment I don't want to wade far into the details. Whole Brain Emulation (WBE) rests on the idea that consciousness emerges from measurable and reproducible physical states of

the brain, which seems to me to be likely but not certain. I don't believe that consciousness consists of any supernatural state, but it seems possible to me that it may depend either on the participation of other parts of the human body (in which case WBE would still be possible, if more complex) and/or on quantum-level phenomena that are probabilistic and thus not fully measurable (in which case WBE would not be possible). Here, as with AIEs, the question that interests me is whether Whole Brain Emulations (WBEs) should be enfranchised morally and politically, assuming that they are indeed possible. As with AIEs, there are several substantial problems in the way.

First, there's the problem of duplication. If a WBE is an upload of a person's consciousness, and the physical person is still alive after the upload, wouldn't giving the WBE a vote in effect be giving the physical person two votes? Presumably over time the two consciousnesses would diverge and come to have different interests and preferences, but at least at first they would be identical, and it's not obvious why duplicating yourself entitles you to additional votes.

Second, there is a set of related problems that revolve around the continuum between a WBE and an AIE. In essence, a WBE is a form of AI, since it is a computer-based sentience. And at least initially it is an AI that is very similar to human beings, so some of the dissimilarities between humans and AIEs aren't relevant. But over time the WBE seems very likely to become less human and more like an AIE. For example, a WBE could reproduce itself infinitely and would effectively be immortal, and presumably those new realities would change the WBE's perspective in ways that would bring it closer to an AIE than to humans (though presumably it would be able to understand humans' perspective from memory, even if it no longer shared it).

Still on the same issue of the continuum between WBE and AIE, consider how a WBE would work. Even if it turns out to be true that consciousness emerges from measurable and reproducible states of the brain, the brain nonetheless is used to receiving enormous amounts of information from the body, and it seems likely that any successful WBE would at least initially need to reproduce that artificially. (I assume that a sudden silence from the body would be profoundly disorienting, perhaps even intolerable.) But what kinds and levels of input should be simulated, and should they be fixed or variable? For example, the brain reacts to stress hormones in the bloodstream. What level of stress hormones should we simulate? Would it be better to have a relatively high level of stress hormones, to improve focus and attention, but perhaps at the risk of rash judgments and impulsivity? Or would it be better to have a relatively low level of stress hormones, to encourage deliberation and reflection, but perhaps at the risk of being unable to treat even serious threats as urgent? Or would it be better to let the levels vary according to context? But in this last case, what is the relevant context? The human stress response is a complex phenomenon that involves the brain, various glands, the blood stream, the autonomic nervous system, the skin, the various senses, and presumably many other somatic processes we don't know about yet. Even aside from the enormous task of figuring out how all of those affect consciousness, how could we effectively simulate them when none of the bodily cues are relevant to what the WBE is doing? The WBE might be thinking about a snake, but there isn't really a snake and the WBE doesn't really have a body to protect, so the fight-or-flight response is situationally inappropriate, though it may be necessary to simulating consciousness and thought. Assuming that we could figure out that thicket of problems, whatever simulation of bodily experience we picked would be guided by some conception of what the WBE's experience is *for*. Should we aim for maximum cognitive speed? greatest creativity? greatest

peace of mind? strongest emotions? best conversation skills? Any such decision would both be arbitrary and would move the WBE further away from being human and closer to being an AIE, since human experience is characterized by constant fluctuations in our experiences, abilities, and immediate goals. Now go through the same process for every hormone, nutrient, neurotransmitter, and somatic feedback pathway that affects the brain, and you have some sense of the scale both of the technical task but also the moral and philosophical problem of whether a WBE is meaningfully human.

Along those same lines, it seems very likely that a WBE would fairly quickly change how it thinks and remembers. For example, we know that some bodily states make it easier or harder to recall things from memory, and conversely we know that some (perhaps all) memories themselves evoke certain bodily states. Imagine that the WBE had a traumatic memory that evoked great fear and stress in the body. If the emulation allows the memory to control the simulated somatic input, then it would be the same memory. But it seems likely that the temptation would be overwhelming to fiddle with one's traumatic memories, to dial back the panic response just a little, and then just a little more. Would a traumatic memory that was no longer associated with a panic response be the "same" memory? It would certainly not have the same behavioral effect, which probably means that it would no longer play the same role in the overall gestalt of one's personality. It's hard to see how it would be the same memory after such changes.

As many studies have shown, human memory is much more fallible and much less objective than we think it is. Memories get edited retrospectively, forgotten, merged with other memories, fabricated completely, and so on. As with emotions and somatic responses, it seems likely that the temptation to try to improve on these processes would be overwhelming. Imagine

for the moment that it could be done, and the WBE's memory could be changed to be more objective, more reliable, and less influenced by transient neuro-physical states. A WBE whose memory worked that way would not be very much like a human being, but would rather be much more like an AIE. It might have perfect recall, but its memories would not be part of a continuing narrative of identity and purpose (which the malleability of our memories supports).

And let's not forget about forgetting. WBE enthusiasts frequently point to the increased possibilities for learning that WBE might enjoy, since presumably it could receive information more or less directly through the computer emulation. (If we know how information is stored in the brain, and we can emulate that, then presumably we can insert new information and/or remove existing information.) That seems possible in principle, but also appears to overlook the tremendous importance of forgetting. What would it be like to be a WBE with twice as much information available in memory? Would it be like being twice as smart as you are now, or would it be like A.R. Luria's mnemonist, who had astonishing powers of recall but was plagued by irrelevant and unwanted memories interfering with his ability to think and act in the world, and who was desperate to be able to forget? Either way, the WBE would be profoundly different from ordinary human beings.

As I have suggested above, the political and moral issues surrounding WBEs turn on the question of whether the WBE is more like a human being or more like an AIE. To the extent that it is more like a human, the normative issues are easier to resolve. To the extent that it is more like an AIE, the normative issues grow increasingly difficult, and increasingly similar to the issues faced by true AIEs.

Conclusion

Obviously, technology exists on a very large spectrum, and the likely impacts of various technologies on citizenship will differ considerably. At the relatively tame and tangible end of the spectrum, we can see that new media and recent developments in computers and devices could help us both recognize civic-minded behavior in a wider range of arenas than we have historically, and envision ways that government could change to encourage greater democratic engagement. Yelping is OK, it is even possibly a citizenship act, and we really could make government a little more like Yelp to our mutual benefit. Robots represent a borderline technology—they're fine as long as they're doing useful work, but potentially problematic when people begin to form emotional relationships with them that ultimately impact the relationships that people have with each other, or when they displace labor in a way that society isn't prepared to manage. When we turn to more speculative technological developments, such as artificial intelligence or whole mind emulation, things seem both more morally urgent and more practically problematic. If we are confronted by machines that are sentient, or human sentiences that have merged with machines, we will have to grapple with the political and moral questions of whether to enfranchise such beings and whether we can coexist with them. As I have suggested above, there are good reasons to think that we would have to treat them as moral equals to avoid committing a great moral wrong, but also that we could not successfully coexist in a single polity, or perhaps even on a single planet. Yet it is a truism that anything that can be done will be done, so either we should earnestly hope that AI and WBE are impossible and will never be achieved, or we should try to figure out innovative ways to solve the problems I've identified above, or we should start packing our bags.

## **Works Cited**



- Driskell, Jay. "An Atlas of Self-Reliance: The Negro Motorist's Green Book (1937-1964)." *O Say Can You See? Stories from the National Museum of American History*. July 30, 2015 2015. Web.
- Kaczynski, Theodore (The Unabomber). "Industrial Society and Its Future (the Unabomber Manifesto)." *The Washington Post* September 22, 1995 1995. Print.
- Lerner, Josh. *Making Democracy Fun: How Game Design Can Empower Citizens and Transform Politics*. Cambridge, Mass.: MIT Press, 2014. Print.
- Luongo, Michael T. "Calvin A. Ramsey on How Black Travel Has Changed since Jim Crow." *The New York Times* July 14 2015, sec. Travel. Print.
- Manjoo, Farhad. "A Plan in Case Robots Take the Jobs: Give Everyone a Paycheck." *The New York Times* March 3, 2016 2016. Print.
- Newsom, Gavin Christopher Dickey Lisa. *Citizenville : How to Take the Town Square Digital and Reinvent Government*. New York: Penguin Press, 2013. Print.
- Schomburg Center for Research in Black Culture, Jean Blackwell Hutson Research and Reference Division, The New York Public Library. "The Negro Motorist Green Book: 1937." 1937. Web.

## Notes

---

<sup>1</sup> As of the end of 2015, Yelp was attracting 86 million unique visitors per month, and those visitors had collectively written a total of 95 million reviews. (<https://www.yelp.com/factsheet>; accessed 4 March 2016)

<sup>2</sup> (Luongo)

<sup>3</sup> (Driskell)

<sup>4</sup> (Schomburg Center for Research in Black Culture)