

DAOs, Democracy and Governance

by Ralph C. Merkle, merkle@merkle.com

Version 1.9, May 31st 2016

FOR REVIEW AND COMMENT

Abstract

Democracies are typically seen as governments which call upon the governed to make the major decisions of government: who shall lead, what policies to follow, what laws to enact. In all these matters democracies call upon ordinary citizens to make complex decisions with eventful consequences.

We do not call upon ordinary untrained citizens to perform surgery, fly airplanes, design computers, or carry out the other myriad tasks needed to keep society functioning, what makes governance different?

The problem is readily understood: if we give governance to “experts” they will make decisions in their own best interests, not in the best interests of us all. As we have seen too often in the past, this leads to enrichment of a small elite and the enslavement or worse of the vast majority. Can we take advantage of the expertise of the best and brightest, while insulating the system from attempts they might make to gain control?

Modern research into “the wisdom of crowds” provides new insights into how to combine the expertise of all participants without handing over control to “experts”. Combined with research on Decentralized Autonomous Organizations (DAOs), this allows us to design a new form of democracy which is more stable, less prone to erratic behavior, better able to meet the needs of its citizens, and which better uses the expertise of all its citizens to make high-quality decisions.

We call this new form of democracy a DAO Democracy.

Cite as: Merkle, R. (2016) DAOs, Democracy and Governance. Cryonics Magazine, July- August, Vol 37:4; Alcor, www.alcor.org

Contents

Abstract.....	1
Introduction	3
The problems with voting	4
The Ideal Democracy.....	6
Governing by prediction market.....	6
DCW: Democratic Collective Welfare	7
Trading Democratic Collective Welfare: an example	9
Membership.....	9
Making it reliable	11
DAO's as living entities.....	12
Replacing Congress	13
Adopting a Bill	14
Self-improving.....	15
Electing the President	15
The Judiciary	16
Other Aspects of National Government	16
Conflict	17
The End of the World	17
Darwin and the DAO's.....	18
Protecting the Ballot Box	21
Voting Redux.....	23
Deployment	23
Summary	25
A Few Characteristics of self-improving DAO Democracies	25
Conclusion.....	27
Acknowledgements.....	27

Introduction

Democracy, n. a system of government in which power is vested in the people.

Democracy, n. an asylum run by the inmates.¹

“... if we want a better politics, it’s not enough to just change a Congressman or a Senator or even a President; we have to change the system...”²

Our lengthening lifespans, the growing lethality of our weapons systems, the coming development of artificial intelligence,³ and the ever more rapid developments in technology make governance ever more important. Failure to develop a system of governance able to respond appropriately and reliably in a timely fashion to the ever quickening pace of change could have major consequences, not the least of which might be to cut short our own lives.

Troubling examples of the failure of governance in all of our major institutions are distressingly easy to find.

War is perhaps the most obvious example where poor governance costs us dearly. Consider that in the short span of 239 years a relatively well run and arguably peaceful democracy, the United States, has been involved in 104 wars.⁴ It is hard to escape the thought that better governance might have significantly reduced both the human and economic toll of many of these wars, and perhaps avoided some of them altogether.

A brief perusal of the news reveals many other examples where improved governance would save lives and reduce human suffering.

The 1991 collapse of the former Soviet Union showed clearly its leaders did not understand how to keep an existing system of governance from collapsing, one of the most basic aspects of governance.

The August 5, 2011 Standard & Poors downgrade of the United States credit rating from AAA to AA+ was a direct consequence of poor governance.⁵ We don’t understand how to prevent major systems of governance from making choices that are obviously and grossly suboptimal. Brinkmanship rewards the useful skill of creating near disasters.

¹ <http://www.davekrieger.net/Glossary/D.shtml>

² President Obama, State of the Union Address, January 12th, 2016.

³ *Superintelligence: Paths, Dangers, Strategies*, by Nick Bostrom, 2014, Oxford University Press.

⁴ To name but a few: the War of 1812, the Mexican-American War, the American Civil War, the Spanish-American War, World War I, World War II, the Korean War, the Bay of Pigs Invasion, the Vietnam War, the Invasion of Grenada, the Invasion of Panama, the Gulf War, the War in Afghanistan, and the Iraq War. In addition, we have had the War on Drugs, the War on Cancer, the War on Poverty and the War on Terror. See the *List of wars involving the United States*, https://en.wikipedia.org/wiki/List_of_wars_involving_the_United_States

⁵ "The political brinkmanship of recent months highlights what we see as America's governance and policymaking becoming less stable, less effective, and less predictable than what we previously believed. The statutory debt ceiling and the threat of default have become political bargaining chips in the debate over fiscal policy." https://en.wikipedia.org/wiki/United_States_federal_government_credit-rating_downgrades

Many countries in the world have little to no governance.⁶ For these countries, even the most basic ability to govern would be an improvement.

In most countries, governance exists but with remarkable lapses. South African President Thabo Mbeki thought AIDS wasn't caused just by a virus, and his health minister, Manto Tshabalala-Msimang, advocated a diet of garlic, olive oil and lemon to cure the disease. Until their removal in 2008, AIDS spread almost unchecked, killing at least 300,000 and making later containment efforts much more difficult.⁷

While some governments might be marginally better, over spans measured in multiple centuries and longer, no existing human institution seems capable of providing the kind of stability and consistently sound decision making that seems required if we are to enjoy long term global security both for ourselves and humanity as a whole, particularly when we consider the exponential advances we are making in technology⁸ and the multi-century and longer lifespans that we expect to be dealing with.⁹

Governance of nations, of companies, of nonprofit and non-governmental organizations (NPOs and NGOs), indeed of any organization, has been subject to major lapses and failures. Even in the best of cases, large and apparently stable and well run organizations suffer from inertia and an inability to respond to new ideas, and seldom maintain their ranking in perpetuity. Organizations that have survived more than a century are relatively rare, have invariably fluctuated in the quality of their governance, and needed a dose of good luck to survive. Even the largest and most stable organizations can't guarantee their own existence, let alone the quality of their governance, for the coming century.

The problems with voting

Our existing democracy is based on voting, yet voting has major problems.¹⁰

First, voters have little incentive to vote at all. In a typical election, the probability is very high that a given vote will have exactly no impact on the outcome. A rational economic analysis would assign an extremely small positive value to voting, even assuming the voter had reason to believe that influencing the outcome would have a discernable positive impact on their life. The time spent voting would exceed the value obtained from the vote. Economically rational voters should not vote.^{11,12}

Second, the voter would have to spend significant time and effort analyzing the various candidates and propositions to determine which, if any, offered any benefit from the voter's perspective. Rationally, this makes voting an even less desirable activity. To the extent social pressure is brought to bear to increase

⁶ The Fragile States Index 2014, <http://library.fundforpeace.org/library/cfsir1423-fragilestatesindex2014-06d.pdf>

⁷ https://en.wikipedia.org/wiki/HIV/AIDS_in_South_Africa#AIDS_denialism_in_South_Africa

⁸ The Singularity is Near, by Ray Kurzweil, Penguin (2006)

⁹ Nanomedicine, Volume I: Basic Capabilities, by Robert A. Freitas Jr., Landes Bioscience (1999)

¹⁰ *The Myth of the Rational Voter: Why Democracies Choose Bad Policies* – (2008) by Bryan Caplan, Princeton University Press. <http://www.amazon.com/Myth-Rational-Voter-Democracies-Policies/dp/0691138737/>

¹¹ *Your Vote Doesn't Count: Why (almost) everyone should stay home on Election Day*, by Katherine Mangu-Ward, November 2012 Reason <http://reason.com/archives/2012/10/03/your-vote-doesnt-count>

¹² *The Empirical Frequency of a Pivotal Vote*, by Casey B. Mulligan, Charles G. Hunter, NBER Working Paper No. 8590, November 2001. <http://www.nber.org/papers/w8590>

“turn out” and cause marginally motivated individuals to vote, their optimal strategy is to acquiesce, make a show of voting, but spend little to no time actually determining how to cast their vote. Such random voting behavior, or worse, voting based on predictable but superficial and incorrect “first impressions”, will result in decision making that is sub-optimal. The cost of determining how best to vote is high, even assuming the voter had accurate information. Again, economically rational voters should not vote.

Third, those spending substantial sums on influencing the voter are providing systematic misinformation intended to cause the voter to select this or that candidate, or vote for this or that proposition, often in direct opposition to the voter’s own interests. If the voter is deceived by one of these misinformation campaigns (which can be very sophisticated) then voting could easily have a very small negative expected value to the voter (instead of a very small positive expected value), and produce systematically poor decisions for society. In any event, unless the voter is prepared to spend even more time analyzing the publicly available information, their vote is unlikely to produce the desired result.

Fourth, candidates are likely to behave, in office, in ways that were not anticipated prior to being elected. Propositions are subject to judicial challenges and other post-election influences that often make the result less predictable than the voter might desire. Not only is voting not worth the time spent on it, not only is it debatable that you can untangle the misinformation, not only is it unlikely that you’ll get what you voted for, in the end you don’t even get what the majority voted for.

Fifth, voting is an activity in which all participate equally, while at the same time it is well known that half the population is below average. We can, at best, expect such a process to produce mediocre results. Voting can only be justified by arguing that the alternatives are worse than mediocre.

From a societal perspective we are expending substantial resources to ensure that unmotivated, unqualified, misinformed voters make societally critical decisions, which are then often ignored. We celebrate this as the pinnacle of modern governance, and shed blood to defend it.

We manage to hold voting in such high regard by comparing it with, for example, absolute dictatorships that slaughter their own population. To paraphrase the old saying: voting is the worst way to make a decision, except for all the rest. But is this really true? Yes, voting is better than Pol Pot’s genocidal regime, or Mao Zedong’s “Great Leap Forward”, which killed between 18 million and 45 million Chinese.¹³ But this is a low bar. Are we content with a system whose major appeal is that it’s better than genocide?

With so many obvious shortcomings, it would seem more than time to put serious thought into how we might produce a better system than democracies based on voting. While it’s true that worse exists, it doesn’t take much imagination to believe that better might be possible.

¹³ https://en.wikipedia.org/wiki/Mao_Zedong, https://en.wikipedia.org/wiki/Pol_Pot

The Ideal Democracy

We might ask: “What is the ideal democracy?”

The phrase we hear most often is that democracies should reflect “the will of the people”. What does this mean? That the goals and interests of the government should be the summation of the goals and interests of the governed?

Consider taking that phrase literally, rather than metaphorically. The goals and interests of the government should be the summation of the goals and interests of the governed. This would imply that the goals and interests of each individual would be given equal weight, and the summation of these goals and interests would determine the global course of action. But how might this abstract concept be rendered specific in a meaningful fashion?

Governing by prediction market

Futarchy, proposed by Robin Hanson, is a proposal to govern by prediction markets.¹⁴ The proposal seems like an excellent approach for improving upon existing democratic forms of governance.

The general concept is to (i) aggregate knowledge from across a community of people, using a method that’s known to be effective: prediction markets.^{15,16} (ii) Use that aggregation method to directly select the actions taken by government. The goals (the “national welfare” to use Hanson’s terminology) are still chosen by voting in futarchy. We discuss later how to eliminate voting entirely.

Governments take action by selecting among bills that are presented to a legislative body for passage. In Futarchy, the bills to be passed are selected by the prediction market, rather than by votes cast by the legislature or the population as a whole. Technically speaking, prediction markets predict rather than select. The first question we have to answer is “what are the prediction markets predicting?” In futarchy, as proposed by Hanson, the prediction markets predict the “national welfare”, as defined by the legislature:

While national welfare could be anything the legislatures chooses, it helps to see that reasonable options exist to choose from. For example, a reasonable initial definition of national welfare could augment current measures of national consumption or product (i.e., GDP) with simple measures of health, leisure, happiness, and the environment.

We choose to be more specific about the definition of what we shall call the “collective welfare”, for the very simple reason that “voting on values” retains the dubious voting mechanism as a core component of futarchy.

¹⁴ *Shall We Vote on Values, But Bet on Beliefs?*, by Robin Hanson, <http://mason.gmu.edu/~rhanson/futarchy2013.pdf>

¹⁵ FORECASTING ELECTIONS COMPARING PREDICTION MARKETS, POLLS, AND THEIR BIASES, by David Rothschild, Public Opinion Quarterly, Vol. 73, No. 5 2009, pp. 895–916. <http://researchdmr.com/RothschildPOQ2009>

¹⁶ Prediction market accuracy in the long run, Joyce E. Berg et al., International Journal of Forecasting, Volume 24, Issue 2, April–June 2008, Pages 285–300. <http://www.sciencedirect.com/science/article/pii/S0169207008000320>

If this is to be a democracy, and if all citizens are to be equal, then all citizens must have an equal say in determining the collective welfare. It is difficult to see how we could choose otherwise. Shall some citizens receive a greater weight than others? That hardly seems like a democracy. Shall each citizen's welfare be evaluated in a manner that the citizen cannot control? Who, or what, can claim greater right than the citizen to evaluate their own welfare?

We define the procedure for measuring democratic collective welfare below, and leave the rather philosophical question of what this democratic collective welfare "really means" for others. We shall simply observe that it clearly treats all citizens equally, and clearly gives each citizen the ability to influence the democratic collective welfare either positively or negatively.

If we are to have a collective welfare, that is, if we are to select a single number to assess the state of our entire society, and if we are to have a democracy in which each citizen's individual welfare has equal weight with that of all other citizens, and if we agree that each citizen should have the right to determine their own contribution to that collective welfare, that leaves us with little room to vary how we compute the democratic collective welfare.

DCW: Democratic Collective Welfare

Annually, all citizens are asked to rank the year just passed between 0 and 1 (inclusive). If you wish, you can think of this as a poll of each citizen's individual welfare, where 0 means the welfare of the citizen that year was the worst possible, and 1 is the best possible. This scale provides no intellectual feedback on whether this or that person should be elected, or which bill should be adopted, or what policy is best: it is intended to provide information about one person's state of satisfaction with the year that has just passed, and each individual citizen selects whatever value they please.

Summed over all citizens and divided by the number of citizens,¹⁷ this gives us an annual numerical metric between 0 and 1 inclusive, or a series of values each one of which summarizes the annual collective welfare of the entire populace for each year. An appropriately weighted sum of annual collective welfares, also extending indefinitely into the future, would then give us a "democratic collective welfare" metric.

More specifically, we can use ACW_i , Annual Collective Welfare in year i , as measured by our direct annual poll, as our base. This year, we would measure ACW_{2016} . Next year, we'd measure ACW_{2017} . The year after, we'd measure ACW_{2018} . And so on. We then define DCW_i , the Democratic Collective Welfare in year i , as 5% of $ACW_i + 95\%$ DCW_{i+1} . DCW_i then gives us a value which depends on future values of ACW_i . Effectively, DCW_i lets us take a look into the future, with progressively declining weights over the next 20 or so years ($1/20 = 0.05$). In some sense, DCW_i lets us look infinitely far into the future, but the

¹⁷ We divide by the number of citizens, n , to produce a number between 0 and 1. This is essential given the way we compute the dollar value of the financial instruments used in the prediction market (that is, as DCW_i times \$1 and $(1 - DCW_i)$ times \$1). It also seems more natural to optimize welfare per citizen, rather than total number of citizens. However, a case can be made that 10,000 modestly happy citizens is better than 1,000 very happy citizens. There is also the question of what to do with the well-known question "are you better off now than you were 4 years ago?"

weights become infinitely small, falling off exponentially the further into the future we go, with a characteristic decay time of ~20 years. That is,

$$DCW_i = 0.05 \times ACW_i + 0.95 \times DCW_{i+1}$$

which can be expanded into:

$$DCW_i = 0.05 \times ACW_i + 0.95^1 \times 0.05 \times ACW_{i+1} + 0.95^2 \times 0.05 \times ACW_{i+2} + 0.95^3 \times 0.05 \times ACW_{i+3} + \dots$$

We provide an example below of how we can trade DCW_i , but the basic idea is that, as time passes, we can convert more and more of DCW_i into ACW_i . In year i , after the annual poll is taken, we can use the equation $DCW_i = 0.05 \times ACW_i + 0.95 \times DCW_{i+1}$ to convert some fraction of DCW_i into ACW_i . The market for ACW_i closes in year i and pays out. In this way, the “indefinite future” market for DCW_i gradually becomes definite and convertible into cash.¹⁸

This kind of “indefinite future” weights the near future more heavily than the far future, and uses a “discount rate” to determine the weighting. We might want to adopt a smaller discount rate, effectively making our prediction market pay more attention to the longer term future, perhaps the next 100 years, rather than the next decade or two. In this way, the “look ahead” of the prediction market can be adjusted. The smaller the discount rate, the longer the look ahead.

Paying attention to any finite period of time is, in some abstract sense, an incorrect strategy. That is, we only pay attention to the next day because we are insufficiently wise to deal with the next week. We only pay attention to the next week because we are unable to deal with the next month. We only pay attention to the next month because we are unable to deal with the next year. And so on. Ultimately, we want to deal with eternity, but we are not yet sufficiently wise. Any non-zero discount rate we choose is, therefore, a concession to our limited mental capacities.¹⁹ Further, the concept of a “discount rate” is itself deeply flawed. Really, we are trying to model our growing uncertainty about the future by using a discount rate. But our uncertainty about the future is not uniform. Sometimes we can make statements about the very far future. To quote Stephen Hawking: “There are certain situations in which we think that we can make reliable predictions, and the future of the universe, on a very large scale, is one of them.”²⁰ At the same time, we can’t predict the roll of a dice even a few seconds into the future. Applying a uniform discount rate to the many events that might occur in our future seems like a heuristic that might be improved upon, if only we were clever enough.

That said, and acknowledging their limitations, at the moment adopting a discount rate seems like at least a plausible heuristic – until a better one comes along. If we think we’re going to be brighter in the future, we could adopt methods that allowed for setting the future discount rate to values progressively

¹⁸ The author’s thanks to Robin Hanson for this clear example.

¹⁹ While a zero discount rate would seem to create problems with infinities, which are difficult to compare, we can always sidestep such problems by using limits: that is, we can ask which policy is better in the limit as our discount rate approaches zero.

²⁰ *The Future of the Universe*, by Stephen W. Hawking, Caltech’s Engineering & Science, Fall 1991.
<http://calteches.library.caltech.edu/664/1/ES55.1.1991.pdf#page=15>

closer to zero. If the discount rate approached zero fast enough, the infinite future would have a significant weight in today's considerations.

DCW_i , Democratic Collective Welfare in year i , is our formalization of the less formal "collective welfare metric". Either one can be replaced with the other. If you want to consider formally what is meant when we discuss evaluation of the collective welfare, we mean DCW_i . If you want an informal description of DCW_i , we mean the collective welfare.

Trading Democratic Collective Welfare: an example

We now give an example of how to buy and sell DCW_i in a prediction market.

First, a trader purchases a pair of conditional bearer bonds from a bank for \$1 in the year 2016. The first says "Pay to bearer \$1 times DCW_{2016} " The second says "Pay to bearer \$1 times $(1-DCW_{2016})$ ". Because DCW_{2016} is between 0 and 1, each of these pays off between 0 dollars and 1 dollar. The two of them together are guaranteed to pay off exactly \$1. The bank, therefore, takes no risk in selling the pair, and simply promises to redeem them once DCW_{2016} is known.

Of course, DCW_{2016} will never be known with perfect accuracy, but as the years go by it will be known with ever greater accuracy. The bank will be happy to exchange a bond that says "Pay to bearer \$1 times DCW_{2016} " for \$0.05 times ACW_{2016} and a bond that says "Pay to bearer \$0.95 times DCW_{2017} ".

The trader who purchased the pair of bearer bonds can now sell the one thought to be less valuable. If the trader thinks DCW_{2016} is actually going to be 0.72, then he will happily sell the bond that says "Pay to bearer \$1 times DCW_{2016} " for \$0.83, netting him $\$0.83 - \$0.72 = \$0.11$. He will then sell the bond that says "Pay to bearer \$1 times $(1-DCW_{2016})$ " for \$0.28. The trader expects to make $\$0.83 + \$0.28 = \$1.11$ for his \$1 purchase of the two bonds from the bank.

In brief summary: the bank issues pairs of bonds to traders in exchange for cash. The bank takes on no risk. Traders buy and sell the bonds, establishing a market for them. Traders speculate on the value of the bonds and trade them to make (or lose) money. The market price of the bonds will fluctuate, depending on events. The bank exchanges the bonds that it has issued for newer bonds and cash, again taking on no risk.

Society benefits by getting reasonably good estimates of the DCW_i .

While this is a simplified example, it conveys the concepts involved in trading in a prediction market for the collective welfare.

Membership

Another problem with evaluating collective welfare is membership in the collective: who do we add and who do we remove? We consider criteria that are suitable for membership in a nation (in contrast with membership in the local chess club, or a student in a school, or a shareholder in a company).

Traditionally, for democracies, children of members are added, and members remain members until they die. In the United States, birth within the United States confers citizenship, although voting rights

are not conferred until some period of time thereafter. Voting rights can be lost following certain judicial proceedings.

We presumably initialize our system with the existing citizens of some jurisdiction or, alternatively, initialize the system with some set of volunteers. Seasteading comes to mind.

For a democracy, and presumably for many other situations, members of the system remain members until they die, at which point we must decide how to treat their death: in future years, what number between 0 and 1 should be used for them when the Collective Welfare is computed?

One answer is the lowest possible number on the scale: 0. This would make death a negative (undesired) event, which the DAO Democracy would then naturally seek to avoid.

A more correct answer to this question is to ask the person who died. Asking them after they have died would be difficult, but we can certainly ask before they die, and if we can't ask every person, we can at least ask most people how they want their death to be recorded, and seek either an answer from an executor or heir, or at least seek a statistically plausible answer if no better answer is forthcoming. Most people will likely regard their death as negative and pick some value close to 0. Some will regard their death under any circumstances as negative, some might not. Some might even say their death is positive.²¹ Regardless, in a democratic form of government the choice can reasonably be left to the individual.²²

We leave the policy with respect to new members, and to births, to our prediction market. There is already a mechanism for adopting policies (see below): we use this mechanism, asserting that the membership that will be used to decide this case will be the existing membership, and will not include new members. In particular if the prediction market concludes that the collective welfare of existing members will improve in the future by adopting a particular policy with respect to new members, then that policy will be adopted.

It is difficult to see how we could justify forcing the members of a DAO Democracy to adopt a policy with respect to new members, or to accept some new born child as a member, if the prediction market says the collective welfare of existing members would be made worse by adopting such a policy.

Contrariwise, it's equally difficult to see how we could justify refusing to adopt a policy that accepts some person, or a new born child, as a member, if the prediction market says the collective welfare of existing members will be improved by adopting such a policy.

There is an additional possible policy with respect to new members. We might allow an existing member to "adopt" a new member by splitting their own, weighted membership in half and giving half to the

²¹ Note that the death of someone who was suffering is still negative, it's just that their life while suffering might have been even more negative, making the moment of death relatively positive compared with the suffering that immediately preceded it, even though their death is negative compared with most of their life. The person who might regard their death as positive would be the criminal on death row who murdered several innocent people and who wishes, quite sincerely and with good reason, that they'd never been born.

²² Although there is usually a requirement that a person be "of sound mind".

adoptee. This would create two citizens, each with half the “weight” of a regular citizen. If continued, this practice could create a small community which, in its entirety, had a “weight” of only 1 citizen. Yet each step along the path was entirely voluntary by those involved.

Allowing some citizens to be “more equal” than others seems to defeat the purpose of a democracy. Yet this question, too, can be reviewed by the prediction market – and if the prediction market favors a policy that allows such a state of affairs to come into being, and if the citizen in question wants to do it, can one reasonably block its adoption? It seems more comfortable to ask: if the prediction market opposes it, how could one reasonably decide to adopt it?

This raises the delicate question of how far a DAO Democracy should be allowed to go in redefining its own purpose. If we provide too much flexibility in this area, then the DAO Democracy could destroy itself. While it seems obvious that it should not be possible to take citizenship away, it’s less clear whether citizenship can be renounced. If a citizen of their own free will, without coercion, wished to renounce citizenship, and a policy was in place which allowed that particular citizen to renounce their citizenship, then it would seem tolerable. Such a policy would have to be adopted by the usual means, of course. And some might require additional safeguards before citizenship could be discarded, else it might happen for reasons that would not withstand closer inspection.

Other than that, the answer would have to be: no, you cannot renounce your citizenship. If there is no policy in place that allows a citizen to renounce their citizenship, then there is no mutual agreement to dissolve the relationship. While it is entirely possible to conjecture what policy a DAO Democracy might adopt on this point, it seems rather clear that, absent a policy that enabled a citizen to renounce their citizenship, it would not be possible.

Making it reliable

We’d like to do all of this reliably, that is, with suitable computer algorithms that provide suitable security. Fortunately, this problem can be very nicely dealt with by implementing a DAO Democracy using a Decentralized Autonomous Organization (DAO), as suggested by Vitalik Buterin.^{23, 24}

DAOs provide a high degree of reliability and incorruptibility, and can be implemented using Ethereum, which makes the power of the Bitcoin algorithm available in a clean and flexible way. Byzantine Fault Tolerance²⁵ (BFT) provides variant approaches that can be used to provide a high degree of stability in a DAO while avoiding the large computational requirements of proof-of-work.

The use of BFT coupled with a prediction market would seem to offer a computationally less expensive and potentially more reliable system. In essence, the prediction market would be used to evaluate the

²³ *Bootstrapping A Decentralized Autonomous Corporation: Part I*, by Vitalik Buterin on September 19, 2013, Bitcoin Magazine, <https://bitcoinmagazine.com/7050/bootstrapping-a-decentralized-autonomous-corporation-part-i/>

²⁴ *An Introduction to Futarchy*, by Vitalik Buterin, August 21st, 2014, <https://blog.ethereum.org/2014/08/21/introduction-futarchy/>

²⁵ https://en.wikipedia.org/wiki/Byzantine_fault_tolerance

reliability of the servers running the core protocols. The most reliable servers (the “core servers”) would be identified by the prediction market. Thus, the BFT algorithms would know who they could most trust, and would use that information as they updated their state information. The prediction markets themselves would, of course, be maintained in a distributed fashion on the core servers, and so would be incorruptible – unless, of course, the BFT algorithms were corrupted. To corrupt the BFT algorithms would require that a majority (or possibly more) of the core servers become corrupted (depending on the details of the BFT algorithm). However, if the prediction markets are reasonably well subsidized, then it will be extremely difficult to corrupt even a few of the core servers, because the prediction markets will have selected the least corruptible servers from among a large number of candidate servers, all of which are running the core protocols. Anyone who spots any irregularity in any core server will immediately make money in the prediction market, and that server will immediately be dropped from the core servers. This will amount to a network of informants, constantly probing for any little concern or worry about any core server.

Breaking the system requires sneaking past a fully functioning and well financed prediction market that is actively looking for any attack and which is running as a distributed algorithm on a set of core servers that are fully protected from any attack. Any slightest hint of any attack that might actually be successful on any core server will result in its immediate removal from the pool of core servers, and its replacement with any one of a large number of constantly running alternate servers. Any strategy for better protecting the core servers, or for better detecting an attack, will be immediately adopted (thanks to a prediction market specifically aimed at improving security which is constantly evaluating new and better strategies). It will be hard to corrupt even a few core servers, let alone half of them. Yet half of them (and possibly more) would have to be corrupted before it would even be possible to gain control of the system and corrupt the prediction market.

Corrupting half of the core servers would presumably require a very intense attack. The other servers would presumably be noticing an alarmingly high number of core servers that were engaged in suspicious activities, creating a full-blown panic. Presumably, such an intense attack on the core servers would be coordinated with additional political and military attacks. At some point the system would succumb, though presumably only after a much more intense attack than any current system could withstand.

Developing and deploying such a system will likely require more time, as it incorporates a prediction market as an integral component.

DAO's as living entities

What is a DAO? The acronym stands for Decentralized Autonomous Organization. The concept for a DAO is derived from Bitcoin, which might be viewed as the first prototype for a DAO.

While there are many technical descriptions of Bitcoin and the excitement it has created, it is perhaps worthwhile to try and capture this excitement in a way that can be understood by those without a technical background.

Briefly, and non-technically, Bitcoin is the first example of a new form of life. It lives and breathes on the internet. It lives because it can pay people to keep it alive. It lives because it performs a useful service that people will pay it to perform. It lives because anyone, anywhere, can run a copy of its code. It lives because all the running copies are constantly talking to each other. It lives because if any one copy is corrupted it is discarded, quickly and without any fuss or muss. It lives because it is radically transparent: anyone can see its code and see exactly what it does.

It can't be changed. It can't be argued with. It can't be tampered with. It can't be corrupted. It can't be stopped. It can't even be interrupted.

If nuclear war destroyed half of our planet, it would continue to live, uncorrupted. It would continue to offer its services. It would continue to pay people to keep it alive.

The only way to shut it down is to kill every server that hosts it. Which is hard, because a lot of servers host it, in a lot of countries, and a lot of people want to use it.

Realistically, the only way to kill it is to make the service it offers so useless and obsolete that no one wants to use it. So obsolete that no one wants to pay for it. No one wants to host it. Then it will have no money to pay anyone. Then it will starve to death.

But as long as there are people who want to use it, it's *very* hard to kill, or corrupt, or stop, or interrupt.

The technical community was fascinated with this new life form. Not because of what it could do, or had done. But because it was a new life form. Suppose you lived on a sterile planet, and all day long you swam in its sterile oceans and watched the waves on its sterile shores. And one day you saw a single trilobite. And you took it apart. And you realized you could build it. And not only could you build trilobites, you could build any of the strange creatures in the Cambrian Explosion. And you dimly understood that there were things beyond the Cambrian Explosion. And you might be able to build them, too.

Anyone who wants to create their own new digital life form can do so. Like Bitcoin, it will live on the internet. Like Bitcoin, it will survive as long as it does something that people will pay for. Like Bitcoin, there will be no way of killing it. Like Bitcoin, it will be radically transparent. Like Bitcoin, it can't be stopped. Like Bitcoin, it will be able to pay people to do anything people are willing to do in exchange for its cryptocurrency. Unlike Bitcoin, it will follow its own rules, whatever rules were programmed into it when it was created.

You could create a currency exchange, or a domain name service, or a prediction market, or a company.

You could even create a government.

Replacing Congress

In some sense, the protocols being proposed are a replacement for Congress. That is, the evaluation and adoption of bills by Congress seems to fit the presently available mechanisms of a prediction market and a DAO quite well. It also seems that the counterproductive incentive structures built into institutions like

Congress, and the variable quality of the laws that it passes, are the most problematic component of our existing governmental structures. Whether or not this ultimately proves to be true, it seems that, at the moment, a relatively rapidly acting Executive able to engage in complex actions based on assessing a wide range of facts and opinions from multiple sources will be more difficult to replace than a slower acting institution which is aggregating facts and opinions from multiple sources in order to review and evaluate proposals placed before it for adoption or rejection in a less time-urgent fashion.

Adopting a Bill

In a DAO Democracy, how do we propose a bill, and how does it get adopted?

Initially, anyone can propose a bill. It can be submitted at any time. If the prediction market says it has a positive impact on the collective welfare, it is adopted. If not, it is not. If the bill is adopted, it's put into effect on the date proposed in the bill, which is typically the adoption date plus some period of time to allow implementation.

At any time, anyone can propose a new method of adopting a bill. It is evaluated and put into effect using the existing methods.

In the United States, this mechanism would replace Congress. Given the current popularity of Congress, any proposal that proposes to replace it will be given a more than fair hearing.

More specifically, how might we propose a bill?

Let anyone propose a bill. Assume the bill includes, as a necessary component, an effective date (which is the adoption date plus time for implementation), and an implementation strategy, or the like. Bills that fail to include such a provision are ill-formed and rejected. Once submitted, there are two possibilities: either the bill is accepted, or the bill is rejected.

Create two conditional futures on the DCW (Democratic Collective Welfare) of the nation. In one conditional future, the bill is assumed to have passed. In the other conditional future, the bill is assumed to have not passed. In each conditional future, there will be some value for DCW: DCW_{passed} and $DCW_{\text{not-passed}}$. If, for a period of 1 week, $DCW_{\text{passed}} > DCW_{\text{not-passed}}$ then the bill is adopted, otherwise it is not. Following the adoption date, one of the two conditional futures is based on a condition that is false. That future market is terminated and all funds returned to the market participants. The other market, which is based on a conditional which is true, continues. Payouts are made from this market based on the actual outcomes that actually occur.

Notice that participants in the market are rewarded (make money) if they accurately forecast the actual value of DCW. They are not "voting for" any particular outcome. A "biasee" who seeks to "bias" the outcome by using the market to predict an incorrect outcome will become vulnerable to anyone who wants to make a profit by correcting the market inefficiency the biasee is creating. That is, attempts to bias the market are, in essence, attempts to create market inefficiencies. To the extent that market inefficiencies can be removed from the system, bias can also be removed from the system. Deliberate

attempts to bias the system for political reasons would, presumably, be a well-known motive and would be watched for, as they would offer a profit opportunity to anyone who corrected the bias.

Self-improving

The components of this system can themselves be upgraded using the system.

For example, suppose we decided that it would improve the stability of the system if all bills had a mandatory minimum consideration period of three months before they could be adopted. Then we would pass a bill modifying the DAO to include this provision. If the DAO is written in a formal programming language, then the bill would specify the change using the formal programming language, as well as the exact place the change was to be made in the existing code, and the time and place the change was to be made.

Of greater concern are changes to the Democratic Collective Welfare metric. Yet even here, if the conclusion reached by the prediction market is that some modification of the metric will better maximize the original metric, then it is difficult to make a case that such a change should be banned. At the same time, it seems unlikely that a DAO Democracy would make any significant changes to the metric that it is supposed to maximize, as that would cause it to maximize something other than the original metric.

There are certain complexities in changing a running program which can most easily be dealt with by specifying that the system be brought to some stable state (stopped) prior to altering the code. Indeed, modifying the system so that it can be safely further modified might be one of the early modifications made to the system.

Over time, all the components of the DAO are likely to be upgraded using its own mechanisms. The only requirement placed on the initial components is that they work well enough that they can be used to replace themselves, thereby initiating a process of continuous self-improvement (Kaizen).

Given the high requirements for stability, it might at first seem that self-improvement could be relegated to a lower priority. However, as further discussion should make clear, such self-improvement will be critical to the survival of any DAO-based democratic system.

Electing the President

Most democracies have an Executive Officer, such as the President of the United States, typically elected either directly or indirectly by the people. In a DAO Democracy, assuming we retain the Executive Branch in more or less its current form, the simplest approach would be to appoint, as President, that person the prediction market said had the highest positive impact on the collective welfare if appointed as President. The President would serve until the appointment of some other person had a higher positive impact on the collective welfare if appointed as President.

An alternative would be to have a special election for the President, using a custom presidential prediction market to select among the candidates. Ideally, in an election for President we'd like to ask the citizens in the different futures created by electing the different candidates what they thought of

them after their terms were over. If A, B and C were candidates, we'd like to hear the historical judgment from (say) a year after the end of their term. If A ranked 60, B ranked 50 and C ranked 40, then we'd know we should elect A, as A was ranked highest by the prediction market just before he was chosen.

Normally, this isn't possible because we can't, as in some science fiction movies, examine counterfactual histories by letting the universe follow multiple different paths and see what happens if some historical event that never occurred did, in fact, occur.

Using conditional prediction markets, we can do just that. The prediction market will tell us how A would have ranked as President if elected, how B would have ranked as President if elected, and how C would have ranked as President if elected. We can then pick the candidate who would have ranked highest. This system is closest to the existing system, in the sense that there is a fixed point in time when we choose a President, and the selection process is one in which citizens are asked, directly, what they think of the President. It's a bit unusual, in that it still uses prediction markets as the core mechanism for making the key choices, and the citizens are asked what they thought of the President who was actually chosen after he has finished his term, and their choice is used to make (likely rather significant) payments to those investors who accurately predicted how people would evaluate this President after his term was finished.

Note the key financial incentive: large payments flow to those who can accurately forecast how the bulk of the citizens will evaluate the President five years in the future. The evaluation is carried out a year after the President's term is finished, and the President's term normally runs four years. Those people who accurately forecast what the vast majority of citizens would think of the President after his term of office had been over for a year will be given a significant financial payoff.

There are many variations on this approach which could be used to elect the President, or any other officer. Which of these variations is best we leave to the prediction market underlying the DAO Democracy to figure out.

The Judiciary

Similar methods could be adopted to appoint the members of the Supreme Court.

So long as decisions can be rendered into one of a few simple formats, such as "yes/no", "a/b/c/d", linear functions or the like, then the use of prediction markets to render decisions should be feasible. However, generating lengthy written decisions by prediction markets appears difficult.

Other Aspects of National Government

Covering other aspects of national government by algorithmic methods using prediction markets and DAOs will require further research and thought. However, many of the problems appear to be tractable and seem to have reasonable prospects for a favorable resolution.

Conflict

A well-known problem with democracies is the tyranny of the majority. This becomes particularly acute when a democracy attempts to unite two groups that hate each other. Each group uses the democracy to weaken the other, kill its members, and in some cases to wipe them out. The primary flaw here is that democracies give “power to the people”, which the people promptly use to kill each other.

It seems that, sometimes, the best solution is to separate the two groups and enforce an armed truce. Democracies seem poorly equipped to carry out this strategy.

What about a DAO Democracy?

There seems no reason to believe a DAO Democracy would fail to recognize the lethality of the situation, nor would it likely assist one group in wiping out the other if there was any way of keeping both groups alive. If there were a pragmatic solution that minimized loss of life, a DAO Democracy would adopt it without hesitation. With a collective welfare metric that placed a high value on human life, a DAO Democracy would be more than willing to separate two warring groups into geographically separate regions if that would reduce deaths. Having once adopted such a strategy, it would be likely to back it up with overwhelming force to minimize the need for actual use of that force.

If the prediction market said such a strategy resulted in a better outcome than alternative strategies, a DAO Democracy wouldn't blink before adopting it.

The End of the World

A problem that might occur with prediction markets and governmental policies might be described as the End of the World problem. Suppose we are evaluating a policy that is very good, but has the small problem that it has a 20% probability of causing the end of the world. Normally, if a prediction market is used to evaluate such a policy, it will decide that it's a wonderful policy because no one will buy a financial instrument that only pays off if they're dead.

To make it more concrete, consider a typical prediction market that is working with two outcomes: either an event occurs, or it does not. A financial institution, such as a bank, issues pairs of bearer bonds, one of which says “This bond pays \$1 on December 31st 2016 if the event has occurred.” The other says “This bond pays \$1 on December 31st 2016 if the event has not occurred.” As only one of the two bonds can pay off, and as the bank issues the bonds in pairs, from the bank's point of view it is simply issuing bonds that pay off on December 31st 2016. From the point of view of the speculator who buys the pair of bonds from the bank, who thinks he knows that the event *will* occur, and who keeps the bond that he knows will pay off and sells the bond that he knows is worthless to some poor fool who will pay him good money for a worthless scrap of paper, this is an easy way to make money from the ignorant.

Prediction markets allow informed individuals to make money from less informed individuals, and the public gets an informed estimate of the probability of the event. If the price of “This bond pays \$1 on December 31st 2016 if the event has occurred” settles at \$0.63, then the prediction market has forecast a 63% probability that the event will occur.

Unfortunately, if the “event” in question is the end of the world, this mechanism fails to work. Who will buy a bond that says “This bond pays \$1 on December 31st 2016 if the world has ended”? And why not pay \$1 for a bond that says “This bond pays \$1 on December 31st 2016 if the world has not ended”?

While amusing and harmless if we’re talking about bonds, the result can be disastrous if we’re talking about a DAO Democracy adopting policies that might actually cause the end of the world if adopted.

Suppose a DAO Democracy is considering whether to build the Large Hadron Collider (LHC), and suppose (for the moment) that there actually was a consensus in the physics community that there was a 20% risk that turning on the LHC would end the world. How do we discover this risk? We can’t simply have a prediction market based on DCW, as the previous example shows the prediction market will simply ignore the 20% risk that the world will end.²⁶

While there might be better solutions, one method is to fall back on simpler techniques: we could have a committee of qualified experts consider the matter and reach a conclusion by anonymous voting one month before turning on the LHC (or the proposed scheduled completion date of the LHC if the project is not funded). This lets us create a prediction market at the time funding of the project is being considered. This market can be used to determine the probability that the LHC would destroy the world. This probability can then be used in evaluating whether or not to fund the construction of the LHC.

Further research on this problem seems warranted. Fortunately, there aren’t many policies that might destroy the world, making special-case handling feasible if a more satisfactory general mechanism can’t be found.

Darwin and the DAO’s

While we have considered DAO’s as a base for a stable democracy, it is apparent that they can serve as a base for any computation.²⁷ Further, given the persistence of a successful DAO (the DAO for Bitcoin has persisted since January 3rd, 2009), it seems inevitable that Darwinian selection will be applied quite vigorously to an ever growing population of DAO’s, all of which are carrying out various useful, or at least DAO-survival-enhancing, functions.

Which raises the obvious question: what enhances the survival probability of a DAO? Today, a DAO survives if it performs some function that causes people to support it. Our economic and digital systems sometimes hide this fact rather well, but it is still the case that people make the wheels of our civilization turn. A DAO that performs some service that some subset of humanity wants, and which cannot be done better in some other way, or which cannot be done better by some other DAO, will be richly rewarded. Those that can’t, will die.

²⁶ Actually, it won’t. There’s likely to be a big stink if 20% of the world’s physicists think the world will end if the LHC is built, and the DAO Democracy goes ahead and builds it anyway. The end of the world might not be noticed, but the political scandal about risking the end of the world almost certainly would be noticed.

²⁷ Blockchain Thinking: The Brain as a DAC (Decentralized Autonomous Organization) by Melanie Swan, Contemporary Philosophy MA Candidate, Kingston University London.
http://www.melanieswan.com/documents/BlockchainThinking_SWAN.pdf

DAO's that incorporate prediction markets and seek to maximize their own profits are obvious, as are self-improving DAO's. The advantage DAO's offer over regular corporations or organizations is a rather radical transparency and incorruptibility.

With Bitcoin, you know what you're getting – or at least, if you're technically expert enough to understand what Bitcoin is doing, you know what you're getting. As Mt. Gox demonstrated, extending an interface that is trustworthy, convenient, and easy to understand remains an important requirement of a system that seeks truly widespread acceptance. A system that allows unproven representatives to mediate between a reliable core and consumers can be badly tarnished when those representatives prove less trustworthy than the core.

Still, if the requirement is a global system to carry out a function with maximal transparency, maximal trust, minimal risk of corruption, and minimal risk of disruption than a DAO seems tailor made for the purpose.

It's reasonable to expect absolutely merciless competition for those functions where a DAO seems suitable.

Hence the need for a DAO with pretensions of serving as a base for a democratic government to be fundamentally and radically self-improving. Anything less will result in a system that will, at some point, be left behind.

The one thing about a democracy that's difficult to "improve" is the collective welfare metric at its heart. This, of course, raises a question of great interest: can a DAO Democracy survive the coming Darwinian competition?

Phrased another way, will a self-improving DAO that has, as its metric, something other than the collective welfare of its citizens, be able to outcompete a similar DAO that is "burdened" with a collective welfare metric? This question might become acutely interesting to citizens of a DAO Democracy at some point in the not-too-distant future.

Arguably, most people would be more willing to support a DAO Democracy than some other form of DAO, as a DAO Democracy obviously and transparently seeks the collective welfare of all its citizens.

There are two reasons a person might choose some other DAO.

First, they might not be a citizen. This is an obvious argument for a DAO Democracy that includes all humans.²⁸

Second, they might support a DAO that preferentially favors them over others. For example, a DAO might seek the welfare of its stockholders, and you might be a major stockholder.

²⁸ A single world-spanning governmental organization is sometimes referred to as a "singleton". This subject raises significant controversy, as an unstable singleton would be highly undesirable.
<http://www.nickbostrom.com/fut/singleton.html>

The question of interest then becomes whether a DAO that seeks to maximize the welfare of its stockholders can defeat a DAO Democracy, or whether the two would reach some sort of mutual agreement. Today, corporations generally abide by the laws that nations create (although the power of larger corporations and smaller nations sometimes overlaps).

Would a DAO Democracy, particularly one which included all humans, dominate other DAO's? While this outcome seems likely, it does not seem a priori inevitable. Indeed, a DAO that maximized its stockholder's welfare might be indistinguishable from a DAO Democracy if every human owned one share of stock, and if the bylaws were appropriately chosen.

This suggests alternatives to a purely egalitarian system are possible. One can imagine a DAO "Democracy" in which there were two or more classes of citizens. Higher classes of citizenship could be awarded based on good behavior, contributions to the democracy (either financial or non-financial), or on some other basis. Might there exist some type of DAO "Democracy" which had well defined classes of citizenship which were conferred on some basis that was widely supported?

In some sense, we already have this: age. All democracies limit participation based on the age of citizens, with some minimum age being required before citizens can vote, drink, drive, or engage in other specific activities. Most societies have licensing requirements for participation in many activities. These have, at times, included requirements for voting, though these have often been used to limit participation by disfavored groups rather than for their ostensible purpose.

For a DAO Democracy, however, exclusion of some citizens from the metric makes less sense than excluding citizens from voting in a conventional democracy. While it is unreasonable to expect a five year old child to make an intelligent decision about which candidate to vote for in a democratic election (the requirement for participation in today's democracies) it is perfectly reasonable to conclude that a starving five year old child is unhappy. The only requirement for participation as a citizen in a DAO Democracy is that, somehow, the DAO Democracy be able to reasonably determine a number between 0 and 1. The annual rating by a starving five year old is unlikely to be a 1, whatever the specific mechanism that might be adopted to ascertain this fact.

If we don't exclude some citizens, perhaps we should reward some citizens by weighting them more heavily in the core metric? All animals are equal, but some animals are more equal than others?

Using the core metric of a DAO Democracy to reward those who have made contributions seems redundant, for a well-run DAO Democracy will need to maintain a smoothly running economic system with well-defined property rights in order to maintain economic productivity. Such an economic system will automatically reward those who are contributing to the collective welfare without having to make additional changes to the system of governance to reward them yet again. Allowing the rich to change the system would seem to be a side effect of *poor* governance. A well run government would reward those who contributed to the welfare of others (by, among other things, making them rich), but would refrain from giving them further power based only on the fact that they have wealth.

Which brings us back to our initial conclusion: a DAO Democracy has one metric: the democratic collective welfare of its citizens. The welfare of all citizens is considered equally (that is, democracy is *egalitarian*). At some fundamental level, the happiness or hurt of one citizen is considered equal to the happiness or hurt of another.

If we are considering a Darwinian competition among DAO's, it is natural to ask if 1) a DAO Democracy or something similar will ultimately win out, or 2) whether a DAO unburdened by concerns about the well-being of humans will be more competitive. If this is the right lens through which to view the evolution of governance, this subject is very much worth further discussion.

Protecting the Ballot Box

The people who cast the votes decide nothing. The people who count the votes decide everything.²⁹ – Stalin

Most forms of government have certain critical infrastructure which must be protected for the whole to function correctly. In democracies, this critical infrastructure involves voting.

In a DAO Democracy this critical infrastructure is discussed below.

The DAO itself is implemented using the mechanisms now familiar to us from the implementation of Bitcoin, though likely using evolved descendants of these mechanisms.³⁰ Whatever the details, the results are likely to be highly secure and difficult to corrupt.

The computations of ACW_i and DCW_i can be made equally secure by adopting suitable protocols.

The most likely point of attack would be the actual inquiry, from citizens, of their response to the annual poll. Corrupt that, and you control the prediction market. Control the prediction market, and you control the DAO Democracy.

Anyone familiar with psychology will recognize that how you ask people how they feel will greatly influence the answer. So, too, with our Annual Collective Welfare survey: how the survey is conducted, and what instructions are provided, and the surrounding publicity and environment, will all have a great impact on the answer. As the DAO Democracy has a great incentive to arrange matters to maximize the answer, meticulous attention will have to be paid to ensure that no undue bias is introduced into the process. The nightmare scenario is a DAO Democracy that directly stimulates the pleasure centers of citizen's brains, leaving them in a permanent state of bliss, utterly uncaring about anything and always

²⁹ This quote from Joseph Stalin is a loose translation of “Я считаю, что совершенно неважно, кто и как будет в партии голосовать; но вот что чрезвычайно важно, это - кто и как будет считать голоса” which more literally translates as “I consider it completely unimportant who in the party will vote, or how; but what is extraordinarily important is this—who will count the votes, and how.” https://en.wikiquote.org/wiki/Joseph_Stalin

³⁰ For example, instead of using Proof of Work, other mechanisms can be adopted. Various methods based on Byzantine Fault Tolerance, Proof of Stake, or a range of alternatives now under active investigation might be adopted. It also seems likely that a better data structure than a simple linear list can be devised. The bitcoin blockchain is already too long. A structure that is more compact, and from which extraneous material can be more easily removed, such as a tree, would seem desirable.

choosing 1 for their individual welfare. This might, arguably, be nirvana, yet it's not something that most of us would want.

It is here, in the central mechanisms used to evaluate the metric, that it might be necessary to create mechanisms in the DAO Democracy that cannot be changed. Obviously, any such unchanging mechanisms must be introduced with great care, as a mistake here can never be undone. At the same time, allowing excessive flexibility might eventually result in unexpected outcomes after several iterations of self-improvement.

Other failure mechanisms include the chain leading from the citizen to whatever centralized computer tallies the score. There is an extensive literature on electronic voting and related systems which can be drawn on.³¹

Interestingly, it's not clear how a self-improving DAO Democracy would react to a threat to the integrity of the Annual Collective Welfare poll. A threat to the lives or well-being of the citizens would provoke an immediate response. The integrity of the annual polls would be protected only if, as a consequence, it threatened the lives or the well-being of the citizens. If there were no perceived threat to the lives of the citizens, and the great mass of citizens were unconcerned about the integrity of the annual polls (which seems likely) then the response of the DAO Democracy itself might be muted.

This might reasonably be viewed as a flaw in the basic design. The most obvious fix would be to ask the prediction market questions of the general nature "Has the prediction market been corrupted?" and "Has the annual poll of the citizens been corrupted?" and use the answers to these questions as negative metrics of collective welfare. That is, the collective welfare metric needs to be augmented with metrics about core aspects of the integrity of the DAO Democracy itself. Such incorporation would lead the DAO Democracy itself to self-improve its own integrity.

Alternatively, one might view such questions as corrections that need to be added to make a prediction market with a shorter term look ahead function more adequately when compared with a prediction market with a much longer look ahead. Failure to maintain the integrity of the core mechanisms of the DAO Democracy will, in the long run, prove fairly lethal to the welfare of its citizens. However, this might not be as evident in the shorter term. By way of analogy, people born without a sense of pain are more likely to suffer from untreated injuries than the rest of us. Pain is a short term indicator that certain types of damage are going to cause problems, and need to be treated immediately. This short term indicator was developed by a longer term process: evolution.

We can add several types of short term pain to our DAO Democracy to assist its survival. Tampering with the prediction market, or the annual poll, or any of the other core processes of the DAO Democracy should result in an immediate and swift response. As humans, we have a long history and a great deal of experience in these kinds of things.

³¹ While it's clear that such systems can be made secure, it's also clear that many systems in wide use today fall woefully short. Slot machines that handle money are often much more secure than electronic voting machines. See Wikipedia's article for an introduction. https://en.wikipedia.org/wiki/Electronic_voting

More generally, it would be desirable to develop mechanisms for modifying the metric so that long-term concerns and problems can be given greater weight when near term patterns are recognized that commonly lead to long term problems. Whether and how to introduce this kind of modification into the metric is a more complex problem than can be dealt with in this paper, but if we provide adequate flexibility in the core self-improvement mechanisms, then any such capabilities, should they prove valuable in the long-term maximization of the metric, can be added later by the DAO Democracy itself.

Voting Redux

How many of the five problems with voting have we solved? We've made little if any improvement in the first, as it still takes a certain minimum amount of time either to participate in determining the ACW (Annual Collective Welfare) or to vote. Whether the former takes a little less time than the latter is perhaps a matter for debate, as the process of voting, having once entered the voting booth, might take a while if there are many candidates and measures to vote on. The latter, picking a single number between 0 and 1 inclusive, should be simpler, or at least no worse.

We've largely solved the second, as participation in determining the ACW eliminates the need to evaluate issues and candidates. The individual citizen is called upon only to determine whether the year has been good or bad for themselves.

We've solved the third, the need to wade through deceptive misinformation. Determining the ACW asks how you feel, not whether others might be deceiving you.

We've likewise largely solved the fourth, the difficulty in knowing whether candidates will carry through on their promises and whether measures will actually be implemented as they were described, as any candidate or measure will continue to be evaluated, each year, by the ACW, and corrective action will be taken by the DAO Democracy should either not produce the desired results.

Finally, we've solved the fifth, as even citizens who are well below average can decide whether the year has been good or bad to them. A DAO Democracy does not disenfranchise below average citizens. It does not ask citizens to analyze the impact of complex proposals, or to evaluate the motives of possibly deceptive politicians.

How does a DAO Democracy rate with respect to these five problems? It solves, or largely solves, four of them.

Four out of five isn't bad.

Deployment

If we grant that the final result is desirable, how might it be achieved? What are the paths from the world as it exists today to this future, better world where sound governance is the rule and not the exception?

An existing organization might already be taking an annual poll of its members, and if the existing poll asks how satisfied each member is, overall, with the organization (or some equivalent question), then

the ACW_i (Annual Collective Welfare) for year i will already be known. Computing DCW_i (Democratic Collective Welfare) for year i is then easy to do. The next step is to establish a prediction market based on the DCW_i. This prediction market can then be used in an advisory capacity to provide an additional source of information about the bills that the organization is already passing. As this leaves the existing organizational mechanisms in place, and simply adds some commentary about the bills the organization is considering, it should not create organizational resistance.

This state of affairs can then be allowed to persist for some time, while the organization gradually gets used to the idea that the prediction market is providing useful commentary on the bills the organization is passing. If all goes well, the prediction market will gradually become more and more respected because its evaluation of the bills the organization passes will prove to be as accurate as other assessments, if not more so.

Once the organization has become familiar with the prediction market, and has seen how it evaluates the organization's bills, and the general opinion is that it can do a good job, the stage is set for the final adoption of the remaining mechanisms of a DAO Democracy.

Call this "gradual adoption".

A more rapid approach would be to start de novo and simply create an organization based on a DAO Democracy. This might be more suitable if there was no pre-existing organizational structure and no need to move slowly. This approach does require that the code base already be written and debugged. Call this "rapid adoption."

Whether adoption is rapid or gradual, we can ask the question: who is most likely to become an early adopter? Adoption of a new form of governance seems most likely by the young, the idealistic, or the desperate. Which means a plausible beginning is with students, idealists, utopians, bankrupt cities, and lawless states.

Some small group with programming expertise will need to spend the several months necessary to get the first kernel of a system up and running for some limited application: perhaps a student group will implement a DAO Democracy for the student government of a college or university, or perhaps a programmer will implement a DAO Democracy for a Seasteading group.

However it's done, once that first implementation is up and running and seems to work moderately well, some small idealistic or utopian community, or a bankrupt city with nothing to lose, will try it out.

Most of these efforts will fall short. Something won't work, some social or technical factor won't be quite right. Eventually, though, an implementation will meet with some success, and the self-improving capabilities will kick in. The system will get better. The people using it will expand it, others will join, the code base will be copied, others will start to use it, and variation and selection will begin.

Someone in Somalia (or some other ineffectively governed region) will pick up a copy and start using it. The infrastructure required is some computers and some cell phones. It will start to work. Any governance at all would be better than what Somalia has, so something that actually worked would start

to build up a following. The basic mechanism should be adaptable to almost any situation. Give it bills and the prediction market will sort out the ones that produce better results for the democratic collective welfare of whoever has become citizens and adopt them. If it works at all, proposals for improvements will be made, the prediction market will pick the likely winners and adopt them, and the system will get better. As long as there's a base of citizens to start with, and a reason for adding new citizens, it will grow. And grow. And grow.

That, at least, is the hope. The process of adoption is likely to start small and depend entirely on how well the system performs with a few hundred citizens. If it does well, more will join. There are likely to be quite a few implementations, with quite a few specific implementation decisions that will have an enormous impact on success. When a good implementation happens to be deployed in an environment where the social factors just happen to be right, the system can take root and start to self-improve.

The next step? Code it up and try it out. The first implementation should be just enough to be useful for some small, well defined group, and have the potential for unlimited self-improvement.

Summary

We can create a DAO Democracy capable of self-improvement which has unlimited growth potential by modifying futarchy to use an unmodifiable democratic collective welfare metric, adapting it to work as a Decentralized Autonomous Organization, implementing an initial system using simple components (these components including the democratic collective welfare metric, a mechanism for adopting legislation (bills)) and using a built-in prediction market to filter through and adopt proposals for improved components.

The resulting self-improving DAO Democracy should grow into something that is very good. Because it uses the transparent and relatively incorruptible digital technology on which the underlying DAO is based, a self-improving DAO Democracy should be a form of governance that is largely impervious to corruption, bribery, irrational bias, and many of the other ills so common in today's governmental systems.

A Few Characteristics of self-improving DAO Democracies

- 1) Anyone can propose a bill at any time.
- 2) All proposals are reviewed equally, without bias.
- 3) Proposals that improve the general welfare (according to the prediction market's evaluation of the publicly known metric for the general welfare) will be adopted.
- 4) The mechanism by which bills are reviewed is open to all.
- 5) Anyone can investigate and present an argument that a bill will help or hurt the general welfare. The judges of that argument (the investors in the prediction market) will have a strong financial interest in reaching an unbiased and accurate conclusion. If the arguments have merit (and have not already been accounted for) the result will be a swift reevaluation of the bill.

- 6) Anyone who wants a more careful review can subsidize deeper investigation, providing an incentive to investigate and potentially change the existing evaluation. Funding such a deeper review does not, however, bias the resulting review. This is a general property of prediction markets.
- 7) The impact on the general welfare is assessed fairly and without bias.
- 8) Any existing law can be amended or repealed with the same ease with which a new law can be proposed.
- 9) Proposals to improve the mechanisms by which bills are reviewed and adopted can be reviewed and adopted by the same mechanism used to review and adopt bills, leading to a self-improving system.
- 10) It's obvious that everyone's welfare is weighted equally, and that no special consideration, either favorable or unfavorable, is meted out to any individual or group, because the mechanisms underlying the DAO are transparent and open to inspection.
- 11) Citizens are not required to exhibit great intelligence, deep understanding of the political system, or to penetrate deceptive campaign statements, for their interests to be properly weighed by the system. The interests of the dull and stupid are protected with as much vigor as those of the intelligent and involved.
- 12) Groups that hate each other cannot directly vote for policies that would harm or even exterminate the other group.
- 13) The only time this governance process would support "the tyranny of the majority" would be if oppression of some minority actually made the majority better off, and the majority was made sufficiently better off that it outweighed the resulting misery to the minority. Jailing Typhoid Mary might have been such a case.³² In the great majority of cases, however, this metric would result in keeping the peace between citizens by some method other than mass jailings, mass arrests, or civil war.
- 14) Once the expected future deaths a given individual might cause exceeded 1, they would promptly be jailed (or otherwise confined). People would likely be confined well before reaching this point, particularly if confinement was not overly burdensome. While accurate forecasting of future murders would seem unlikely, if such a technology were feasible it would be developed and promptly used by this metric to confine people who were likely to kill someone. On the other hand, the proposed metric (a person's death typically resulting in future individual welfare ratings of 0) would create an aversion to the death penalty.
- 15) Warfare, if it involved deaths of citizens, would be avoided. If it involved deaths of non-citizens, it would be pursued vigorously if it produced benefits to citizens. In this latter case, the only thing preventing war would be the rational expectation that peace would be better for the collective welfare. This might well be the case, but drone warfare might be pursued in those cases that did not create undue outcry by the international community. Unless the proposed metric is extended to the entire

³² Though other methods of ensuring she was not spreading contagion might have been effective.

human population, there is no substantial incentive not to kill or injure non-citizens, provided there is some advantage to be gained and little risk to citizens. The metric, as given and without further modifications, treats non-citizens as having no value. This, of course, creates an argument for extending the metric to the entire human species.

Conclusion

Modern results in distributed algorithm design (particularly including the proposal of Decentralized Autonomous Organizations), combined with modern insights into aggregation of knowledge from dispersed individuals (prediction markets), can be combined into novel forms of governance that are self-improving and should be more robust and resistant to traditional failings and weaknesses than existing forms. Further analysis and small-scale implementations, combined with further research into their effectiveness, seem both warranted and urgently needed, given the range of problems facing humanity today and the more pressing problems anticipated in the future.

Acknowledgements

It is the author's pleasant duty to acknowledge the comments and criticisms that have made the present paper as good as it is. Any faults that remain are, of course, the authors. The author thanks: Nick Bostrom, Vitalik Buterin, Robert A. Freitas Jr., Robin Hanson, Charles Hoskinson, John Oh, and Melanie Swann.