On the Limited Applicability of Liquid Democracy

Ben Armstrong
University of Waterloo
Waterloo, Ontario
ben.armstrong@uwaterloo.ca

Kate Larson
University of Waterloo
Waterloo, Ontario
kate.larson@uwaterloo.ca

ABSTRACT

Liquid democracy is a proxy-voting method in which voters may transitively delegate their votes, allowing others to vote in their place. It is often analysed and used in settings where there is some single, objectively correct outcome that voters collectively aim to select. As a result, much of the previous work on liquid democracy has proposed agent utility functions which depend directly on the accuracy of the delegatee voters. We argue, instead, that a more realistic model would have voters' incentives more closely aligned with the final outcome, rather than the proxy vote cast. We explore this alternative and show that there exists a pure-strategy Nash equilibirum that can be reached by best-response delegations. Furthermore, allowing for vote-delegation always weakly improves the accuracy of the decision-making process compared to having all agents vote directly. We apply this model to a classifier ensemble problem. While our theoretical findings are positive, our empirical results show that the assumption of independence between voters required for theoretical analysis is critical. Once removed, liquid democracy fails to materialize any practical improvement over direct voting.

KEYWORDS

liquid democracy; computational social choice; Nash equilibrium; voting simulation; machine learning

ACM Reference Format:

Ben Armstrong and Kate Larson. 2021. On the Limited Applicability of Liquid Democracy. In *Appears at the 3rd Games, Agents, and Incentives Workshop (GAIW 2021). Held as part of the Workshops at the 20th International Conference on Autonomous Agents and Multiagent Systems., London, UK, May 2021*, IFAAMAS, 7 pages.

1 INTRODUCTION

Liquid democracy is an emerging trend in social choice aimed at allowing less informed voters to benefit from expert knowledge and find a balance between traditional representative democracy and direct democracy. It is a form of proxy voting [17] in which delegation is transitive. Every voter has the option of casting their vote directly or of delegating to another voter. Agents that vote directly are referred to as *gurus* and their vote is given a weight equal to the number of direct and indirect delegations they have received (including their own). While not inherent to liquid democracy it is common that a weighted majority voting rule is used.

First described in 1969 [24] modern technology has allowed liquid democracy to be practically implemented and used with up to thousands of voters [2, 18]. These systems aspire to provide all the benefits of direct democracy by allowing all voters who are well-informed on a particular issue to vote directly, while also – as in representative democracy – giving non-experts the ability to support expert decision-makers. For these reasons many models restrict their focus to a *ground-truth* setting where voters must select a single correct outcome from two (or more) alternatives.

However, a less optimistic view of liquid democracy reveals several potential concerns [4]. Allowing arbitrary delegations gives the possibility of single voters receiving disproportionate power, which has already been observed in practice[21]. Cycles among delegators can occur frequently and there seems to be no clear consensus about how to resolve them with some work restricting delegations to disallow cycles [20], effectively ignoring voters in a cycle [3], or requiring voters provide additional information such as a ranking over possible delegations [7]. As well, there has thus far been relatively little work quantifying the benefits of liquid democracy over other social choice mechanisms.

This paper explores the final issue: whether it is possible to reliably benefit from liquid democracy. We study a basic model of ground-truth voting with two alternatives and provide the first formalization of group accuracy in this setting within liquid democracy. Our theoretical results show that allowing delegation between voters will always weakly improve the accuracy of a group at finding a ground truth. This model is validated experimentally by extending it to a machine learning setting and using classifiers to act as voters. However, the improvement from liquid democracy tends to be relatively small and finding the ideal delegations to achieve it is computationally expensive. We go on to demonstrate that achieving significant gain to group accuracy from simpler delegation strategies is not readily possible. This throws into question whether liquid democracy is actually a good choice for ground-truth elections.

We begin with an exploration of the existing work on liquid democracy, highlighting several existing and unresolved issues. Section 3 introduces our formal problem setting and model describing how voters decide whether or not to delegate and the delegation process. This model is analysed in Section 4 with results showing a guarantee that giving voters the option of delegating can always weakly benefit group accuracy. Section 5 extends our model to fit a real-world setting and runs liquid democracy on classifier ensembles. Experiments in this setting show that reliably gaining any significant benefit from liquid democracy is not feasible. We conclude with a call for future work in this area of social choice to pay more heed to empirical limitations of their theoretical results.

Appears at the 3rd Games, Agents, and Incentives Workshop (GAIW 2021). Held as part of the Workshops at the 20th International Conference on Autonomous Agents and Multiagent Systems., Aziz, Ceppi, Dickerson, Hosseini, Lev, Mattei, McElfresh, Zick (chairs), May 2021, London, UK.

2 RELATED WORK

Delegative democracy has a history tracing back as far as Dogdson's work in 1884 [14] and has since been studied in political science contexts [4, 17]. Recently delegation has gained attention in the artificial intelligence and social choice communities, with a focus on liquid democracy. Modern work in this area tends to involve modifications or clarifications to the basic concept of liquid democracy or analysis of such a system.

A model in which voters aim to identify a shared ground-truth and have some probability of voting correctly is studied by [3] which considers the existence of Nash equilibria but defines a utility function for each voter rather than focusing on group accuracy as a whole. [20] also considers a similar model and shows that under certain assumptions there is no decentralized mechanism that can guarantee delegation will strictly improve accuracy. Some of these assumptions are rebutted by [10] which shows that optimal delegations cannot be easily approximated but can require delegating to less accurate voters.

Several papers aim to address potential problems with liquid democracy. Delegation cycles naturally occur often and may deprive voters of representation; [22] develops a method of resolving delegations which avoids cycles while avoiding cycles is shown to be NP-hard in some settings by [8] who adapt an existing framework to take into account delegation graphs while voting rather than trying to avoid cycles. [5] address the issue of possible dictatorships arising by reducing the weight of delegations based on how far they have traveled while [16] develops a centralized mechanism that aims to minimize the maximum weight given to any voter by allowing voters to specify multiple delegations. [26] aims to understand how power is distributed among voters and what leads to disproportionate accrual of power.

Empirical uses of liquid democracy are studied by a few papers, though they frequently do not draw connections with existing theory or provide significant evaluation of their systems. [21] describes the use of liquid democracy in German elections and shows the existence of "super-voters" with many more delegations than most individuals. Others, such as [6] and [1], describe practical uses of liquid democracy but lack evaluation.

The experimental design of our ensembles is based on [13] which investigates the impact of different voting rules on ensemble accuracy. Likening ensembles of classifiers to voters bears a similarity to the theoretical model of liquid democracy studied in [11], where voters must decide on many binary propositions. They show the existence of a number of problems relating to rationality and suggest as solutions multiple voters that select multiple possible delegates and default votes. To gain insight into pre-existing measures of ensemble diversity we have primarily drawn from two survey papers, [9, 23], which define several group and pairwise measures of diversity.

3 MODEL

In this section we describe the agent voting model that we study. We assume there is a set of alternatives, $A = \{a^*, a^-\}$, where a^* is, objectively, the correct outcome and a^- represents the other alternative.[1] There is a set of n voters, $V = \{1, \dots, n\}$. Each voter receives a noisy, independent signal as to which is the correct outcome, which we refer to as the voter's accuracy. In particular, we say that each voter $i \in V$ has accuracy level $q_i > 0.5$ which is the probability that they view a^* as the best alternative.

Voters have two actions available to them. First, a voter is able to directly cast a vote by indicating their chosen alternative to the voting mechanism. However, we assume that direct voting is costly. That is, if voter i decides to vote directly, then they incur cost $c_i \geq 0$. This cost captures the "effort" of voting. Alternatively, they may *delegate* their vote to another voter.

Definition 3.1 (Delegation). Let $i \in V$ be a voter. A delegation occurs when i selects some $j \in V$ and grants them permission to vote on their behalf. We call voter i the delegator and voter j the delegate.

If voter i delegates to another agent j, then they pass all responsibility of voting to the delegate, j, including ceding all control as to which alternative will be voted for. Furthermore, it is possible for a delegated vote to be further delegated. That is, if voter i delegated to voter j, then voter j may also delegate it (and i's) voting privileges to voter k. We let $d(i)$ be the delegation action of voter i and let $d^*(i)$ represent the recursive application of $d(i)$ until a fixed point is reached. We refer to the final set of active voters as *gurus*.

Definition 3.2 (Guru). Voter $j \in V$ is a guru if $j \in \{d^*(i) | i \in V\}$. We let $G(V)$ denote the set of gurus given V. We say that j is the guru of i if $j = d^*(i)$.

Observe that it is possible that i delegates to itself, and so i is its own guru when $d^*(i) = i$.

Voters always have full knowledge of the current delegation structure and each voter's accuracy. Voters exist on a complete network and may delegate to any voter, unless that delegation would cause a cycle.

Through our analysis we focus on the setting where the following procedure is used to determine delegations:

(1) Initially, $\forall i \in V$, $d(i) = i$.
(2) Proceeding in arbitrary order, each voter chooses a delegate (possibly unchanged from their current delegate).
(3) Step (2) is repeated until a round passes in which no voter alters their delegation.

We assume that delegation decisions happen before any voting actions can be taken. Once the set of gurus has been identified (i.e. $\{d^*(i) | i \in V\}$ has been determined), then all *gurus* submit their ballot to a central authority, which uses weighted majority voting to determine which alternative is selected. The weight, w_j, of a guru $j \in V$ is a function of the number of voters it is *representing* through delegation actions. In particular, for any $j \in V$

$$w_j = \begin{cases} 0 & \text{if } j \notin G(V) \\ |\{v \in V | j = d^*(v)\}| & \text{otherwise} \end{cases}$$

As mentioned earlier, the action of voting is costly, with agent i incurring cost $c_i \geq 0$ is actively voting. We assume, however, that if $w_i = 0$ then the agent does not vote (since they are not a guru) and thus incur no cost.

Next we introduce the utility function that is used by voters to make delegation decisions.

[1] This restriction to two alternatives is common in the literature [3, 10, 20]

3.1 Formulating Utility Functions for Gurus

Previous work used a similar model of liquid democracy where agents had different levels of accuracy [3]. However, they assumed that the underlying utility model each voter depended on the accuracy of their guru, namely $u_{old}(i) = q_{d^*(i)}$. While appropriate for their analysis, we argue that there are many application-domains for which this assumption is not appropriate. Instead, we introduce an alternative utility function which we believe has broader applicability and analyse it to better understand the importance of modelling decisions.

Since voters do not benefit from concealing their accuracy and delegation we assume each voter has full knowledge of the entire problem instance. Voters are all aiming to select the same optimal outcome and we are interested in maximizing the probability of this outcome being selected. Thus, we argue, a utility function that accurately represents the agents' underlying goals is appropriate. We achieve this by giving all voters the same utility: the exact chance that the optimal candidate will be selected. Our utility function is:

$$u_{new}(V,d) = \sum_{l=\frac{n}{2}}^{n} \sum_{S \in G_l} (\prod_{i \in S} q_i \prod_{j \notin S} 1 - q_j)$$

where $G_l = \{\{v_i \in V_l \mid d^*(i) = i\} \mid V_l \subseteq V \text{ and } \sum_{v_i \in V_l} w_i = l\}$ contains all sets of gurus whose combined weights sum to l. Each value in the summation over G_l corresponds to the probability of exactly that set of gurus voting correctly, while all others vote incorrectly. This is repeated for each value of l greater than or equal to half the total number of voters. If voting directly has a non-zero cost, voter i subtracts c_i from u_{new} to calculate its individual utility.

4 EQUILIBRIUM ANALYSIS

We now show that with our utility function, voters making sequential best-response updates to their delegations will always result in a pure strategy Nash equilibrium. This applies to the case where voting directly does not cost anything, and the case where there is some small cost $c_i > 0$ required to act as a guru.

Theorem 4.1. *A pure strategy Nash Equilibrium always exists for the effortless voting scenario.*

Proof. The following procedure will result in a Nash Equilibrium: Begin by having all voters cast their votes directly. Choose voters in arbitrary order and have them select the delegate that will lead to the highest utility (i.e. any connected voter that will not cause a cycle, or themselves). Repeat this until no voters will increase their own utility by changing their delegation.

To confirm that this final state is a Nash Equilibrium, note that all voters have the same utility function. Thus, any best response action taken will never decrease any voter's utility. When there are a finite number of voters this means that the process will eventually stop since there are a finite number of utility improvements that can take place. □

Theorem 4.2. *A pure strategy Nash Equilibrium always exists for the scenario where $c_i > 0$.*

Proof. Say that voter i has utility of $u - c_i$ while voting directly and decides to delegate to voter j, leading to utility of u'. This means $u' > u - c_i$ and there are two cases:

(1) $u' > u$: In this case, all voters delegating to i (as well as all other voters) will have strictly higher utility than before i's delegation to j and the process of following best responses will lead to a NE as above.

(2) $c_i > u - u' > 0$: This means that i has increased their utility but all other voters have had their own utility reduced by an amount up to c_i. There are a number of possibilities here: If j is not a guru, their only actions are to (1) switch delegations which will strictly increase the utilities of all players, or (2) begin voting directly which implies that the new accuracy $u'' > u'$ since $u'' - c_j > u'$ which will also strictly increase the utility of all players.
If j is a guru when they receive i's delegation, their only action is to change delegations. Say j's best response action is to delegate to k and their utility becomes u''. If k directly or indirectly delegates to i this would cause a cycle which is not possible so k must either vote directly, or delegate in such a way that does not cause a cycle. If k currently delegates, the reasoning in the previous paragraph applies. If k is a guru, the reasoning in this paragraph applies. In either case, assuming a finite number of voters, the delegations will reach a fixed point and there will be no more best response actions to take.

□

Every delegation change is guaranteed to increase the utility of the delegator so a clear upper bound on the number of actions required to reach equilibrium is the total number of sets of delegations. This is equivalent to the product of each voter's network degree: $\prod_i^n l_i$ where l_i is i's degree. For our setting, with a fully connected network there may be as many as $O(n^n)$ states. While some of these states may be easily discarded (e.g. it is easy to check if having a single dictator is an equilibrium, delegation cycles are disallowed) reducing this bound more generally proves complex.

However, in practice, we see voters quickly converge to a Nash Equilibrium. Table 1 shows that elections of a tractable size require, on average, very few best-response updates. Table 1 also shows that while utility increases as a result of delegation (in fact, this is guaranteed when voting directly is costless) it tends to increases by a relatively small amount. Note that due to the size of G_l in our utility function evaluating significantly larger voter populations in this manner is not possible. Recent work has found an efficient way of calculating utility when voters all have identical weight (in which case it is equivalent to the Poisson-Binomial function CDF) [19]; extending this result to a weighted setting would be widely useful future research.

5 USING LIQUID DEMOCRACY WITH ENSEMBLES

The model described in Section 3 applies to voters choosing between a single pair of alternatives. To evaluate the efficacy of liquid democracy on real-world tasks we now show how this model maps to a setting where each voter selects between many pairs of alternatives

n	5	11	17
Actions	1.21 ± 0.71	3.18 ± 1.11	4.78 ± 1.43
Util Gain	0.035 ± 0.029	0.028 ± 0.018	0.017 ± 0.011
Start NE	0.17	0.01	0.0

Table 1: Mean and Std Dev of number of actions required to reach Nash Equilibrium and increase in utility between initial and NE states across 100 randomly initialized voter populations for $n = 5, 11$, and 17 voters. The third row shows fraction of simulations in which the initial state was a Nash Equilibrium.

simultaneously. In particular, we draw a parallel between voters selecting an alternative and machine learning classifiers predicting classes on a dataset then present results of several experiments done in this setting.

Consider m pairs of alternatives $A_m = \{(a_1^*, a_1^-), ...(a_m^*, a_m^-)\}$, corresponding to the m *examples* within a two-class dataset. Each of the n classifiers receives noisy information about the correct alternative in each pair from feature values in the dataset. Classifiers have two actions available to them: First, they may directly predict a class for each of the m examples (ie. vote m times). Unlike our original model, there is no cost to direct action. Alternatively, they may delegate to another classifier who will predict (vote) on their behalf for *all* examples. We assume that each individual classifier voting directly correctly classifies at least $\lceil \frac{m}{2} \rceil$ examples.

Our original utility function focused on the accuracy of a group of voters at selecting a single correct alternative. Now we are primarily interested in the total number of examples correctly classified by an ensemble of classifiers. If A_m^{win} refers to the m alternatives from A_m that won each individual election then the accuracy of an ensemble is acc $= \frac{1}{m}|\{a_i \text{ if } a_i = a_i^* \; \forall a_i \in A_m^{\text{win}}\}|$.[2] Our experiments also consider a number of different delegation mechanisms, the explanations of which are deferred to Section 5.2.

Since our focus is now on the accuracy of groups of classifiers, the results from Section 4 no longer directly apply. Rather, we now shift our attention to better understanding the chance that delegations will benefit an ensemble of classifiers. Going forward when we refer to a voter we are referring to a classifier voting as described above, and use the terms interchangeably as best fits the current discussion.

5.1 Chance of Weak Improvement from Delegation

Here we show that the large majority of possible delegations lead to a weak benefit for group accuracy. In particular, we show that from all possible sets of ensembles where each classifier votes directly very few of these guarantee that every possible initial delegation reduces accuracy.

To do this we consider the votes of a set of classifiers with no delegations. These can be treated as an $n \times m$ binary matrix P where p_{ij} is 1 if voter i voted correctly on the j^{th} example and 0 otherwise. Each row must sum to at least $\lceil \frac{m}{2} \rceil$ and any column that sums to $\lceil \frac{n}{2} \rceil$ or more indicates the corresponding example is classified correctly.

An example j is called *pivotal* if $\sum_{i=1}^{n} p_{ij} = \lceil \frac{n}{2} \rceil$. That is, if it is correctly classified by a minimum margin. Any voter that is correct on a pivotal example is said to be a *pivotal voter* on that example, and may be pivotal on several examples. Similarly, if a voter is incorrect on example j where $\sum_{i=1}^{n} p_{ij} = \lceil \frac{n}{2} \rceil - 1$ they are considered an *incorrect pivotal voter*. We can now begin to establish an upper bound on the number of states in which any individual delegation would result in a decrease in accuracy (and thus, well-informed greedy voters would make no delegations).

Theorem 5.1. *If (1) in each pair of classifiers each member in the pair is pivotal on an example that the other voter classifies incorrectly, and (2) there are no incorrect pivotal voters then any single delegation in an ensemble of equally weighted classifiers reduces group accuracy.*

Proof. Say that voter i delegates to voter j and that example s_{ij} is an example on which i is pivotal and j classifies incorrectly. s_{ij} always exists by the first condition. If this delegation occurs s_{ij} will no longer be classified correctly and, since i was not an incorrect pivotal voter by the second condition, j's increased weight will not cause any incorrect examples to become correct. □

Our goal is to show that very few matrices P exist that satisfy the conditions in Theorem 5.1. Our next result can be used directly to give an upper bound on the number of such matrices.

Theorem 5.2. *Any matrix P that satisfies condition* (1) *in Theorem 5.1 with n voters must have at least n pivotal examples.*

Proof. Note that a prerequisite for satisfying (1) is that for each pair of voters there exists a pair of correctly classified examples on which each voter correctly classifies exactly one of the examples.

We proceed with proof by induction. For $n < 3$ it is clear that there are at least n pivotal examples when (1) is satisfied. When $n = 3$, consider a pair of voters and two examples. Each example is correctly classified exactly once; adding a new voter and no new examples while continuing to satisfy the condition is impossible thus there must be at least 3 examples in order to satisfy (1).

Now we assume that any group of $n \geq 3$ voters requires at least n examples and show that (1) cannot continue to be met when we add a new voter and no new examples.

Denote the newly added voter as v_k, the existing voters as $V = \{v_1, v_2, ..., v_n\}$, and the examples required for V to satisfy (1) as $E_V = \{e_1, e_2, ..., e_n\}$. By the induction assumption we know that voters $V' = \{v_1, v_2, ..., v_{n-1}, v_k\}$ require (at least) n examples, $E_{V'}$. If $E_V \cap E_{V'} \neq E_V$ then $V \cup \{k\}$ requires at least $n + 1$ examples.

Otherwise, $E_{V'} = E_V$. Now it must be the case that voters v_k and v_n classify all examples in E_V identically. To see why, consider $V'' = V \setminus \{v_n\}$ and the $n-1$ examples required for V'' to satisfy (1). In order for each example in $E_{V''}$ to remain pivotal any new voter is confined to a particular classification on that example. Thus v_k and v_n must be identical on each example in $E_{V''}$ and when they are both added to V'' two new examples are required in order to satisfy (1), leading to $n + 1$ examples. □

[2] While accuracy is not always the most responsible metric to consider we focus on it here due to the natural parallel with our original model.

Dataset	Examples	Categorical	Numerical	Missing
breast-cancer-w	699	0	9	yes
credit-approval	690	9	6	yes
heart	270	0	13	no
ionosphere	351	34	0	no
kr-vs-kp	3196	0	36	no
spambase	4601	0	57	no

Table 2: Description of datasets.

	n			
m	3	7	11	15
3	5.00e-02	-	-	-
7	6.25e-03	5.61e-03	-	-
11	4.10e-03	5.47e-07	1.33e-04	-
15	3.60e-03	7.36e-08	2.79e-12	7.98e-07

Table 3: Examples of (an upper bound on) the ratio of the number of states in which any delegation is harmful to the total number of possible ensemble states the number of examples (m) and the number of voters (n) increase when there are an equal number of voters and pivotal examples.

5.1.1 Counting Harmful Delegations. Using the limitation from Theorem 5.2 we now show that the number of "harmful" vote matrices P guaranteeing a reduction in accuracy from any single delegation is tiny when compared to the total possible number of such matrices.

First, note that if we assume each voter is accurate on at least $\lceil \frac{m}{2} \rceil$ examples (if they were not they could simply reverse their classifications to improve their accuracy), the total number of unique voters is $v^{\text{unique}} = \sum_{c=\lceil \frac{m}{2} \rceil}^{m} \binom{m}{c}$. Voters in the same ensemble may be identical to each other so the total number of possible ensembles of size n is $\binom{n+v^{\text{unique}}-1}{n}$.

We now consider an upper bound on the number of possible harmful vote matrices. Assume there are n pivotal examples and each voter is correct on $\lceil \frac{n}{2} \rceil$ of them. The number of possible voter distributions on pivotal examples is $v^{\text{pivot}} = \binom{n}{\lceil \frac{n}{2} \rceil}$ allowing up to $\binom{v^{\text{pivot}}}{n}$ sets of voters (chosen without repetition to satisfy (1)). On the remaining $m - n$ examples, each voter must be correct on $\lceil \frac{m}{2} \rceil - \lceil \frac{n}{2} \rceil$ to $m - n$ of them. This gives a total of $v^{\text{remainder}} = \sum_{c=\lceil \frac{m}{2} \rceil - \lceil \frac{n}{2} \rceil}^{m-\lceil \frac{n}{2} \rceil} \binom{m-n}{c}$ options for each remaining voter. Votes on the remaining examples may be repeated across voters so there are $\binom{n+v^{\text{remainder}}-1}{n}$ options here.

Putting this together gives an upper bound of $\binom{n+v^{\text{remainder}}-1}{n}\binom{v^{\text{pivot}}}{n}$ possible states where delegation will always be harmful. Table 3 gives some insight as to the proportion of states in which any single delegation will be harmful, for certain values of m and n and the case when there are an equal number of voters and pivotal examples. As the size of the dataset and number of voters grows the chance that delegation will allow for a benefit grows quickly.

dataset	best	full	diverse	max	random
breast-cancer-w	0.951	0.943	0.944	0.951	0.94
credit-approval	0.87	0.862	0.863	0.868	0.861
heart	0.80	0.762	0.758	0.802	0.771
ionosphere	0.902	0.884	0.889	0.903	0.878
kr-vs-kp	0.921	0.905	0.909	0.909	0.881
spambase	0.883	0.869	0.866	0.884	0.862

Table 4: Accuracy of several types of ensemble on each dataset averaged over 5, 7, and 9 voters with 3 and 5 gurus.

5.2 Experimental Setup

To evaluate the usefulness of liquid democracy on real-world tasks we run experiments using 6 two-class datasets from the UCI Machine Learning Repository [15]. Table 2 contains brief summary information about each dataset, note that examples with missing feature values were left out of training.

We loosely base our experimental setup on that of [13]. As such our ensembles are composed of decision trees using gini and entropy critera with equal likelihood. The maximum depth of each tree was chosen uniformly randomly between 1 and 4. All experiments were implemented in Python 3.6 using the scikit-learn library for all machine learning functions [25]. Initially we also experimented with ensembles that included SVM's and Neural Networks with equal probability as Decision Trees but found no significant difference in performance beyond significantly increased training time.

With our experiments a primary goal was to evaluate a range of conditions and understand when and by how much liquid democracy might improve the accuracy of an ensemble. Thus, rather than allowing voters to decide for themselves whether or not to delegate at any time we use centralised mechanisms to have voters delegate until a pre-set number of gurus, n^{final}, exist. These mechanisms aim to either increase independence of voters or to remove the least accurate voters. All experiments are averaged over 5 trials with different randomly initialized ensembles. In each trial we perform 10-fold cross validation on each ensemble type.

5.2.1 Classifier Diversity. A significant differences between our earlier setting and the use of classifiers as voters is that we cannot guarantee the independence of classifiers being trained on real-world datasets. To ameliorate this issue we aim to increase classifier/voter diversity when deciding upon delegations.

Diversity of classifiers is studied in depth in by [23] where they introduce a number of metrics to measure diversity of both entire ensembles and pairs of classifiers. Our experiments explored a number of diversity metrics, including the disagreement and double-fault measures and variants thereupon, and found the *q-statistic* performed the best. It is defined as follows: For trained classifiers i and j, if N^{11} is the number of samples both predicted correctly, N^{10} is the number i predicted correctly and j incorrectly, etc. Then,

$$Q_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}}$$

5.2.2 Delegation Mechanisms. We now introduce the 5 delegation mechanisms we explore in our experimental results.

(a) 9 voters, 5 gurus					(b) 29 voters, results averaged over 5, 15, and 25 gurus				(c) 49 voters, results averaged over 5, 15, and 25 gurus			
dataset	full	diverse	max	random	full	diverse	max	random	full	diverse	max	random
breast-cancer-w	0.943	0.939	0.951	0.945	0.942	0.947	0.955	0.946	0.946	0.949	0.96	0.943
credit-approval	0.862	0.864	0.865	0.861	0.862	0.863	0.87	0.862	0.863	0.863	0.872	0.861
heart	0.772	0.768	0.785	0.766	0.76	0.767	0.805	0.767	0.767	0.774	0.82	0.762
ionosphere	0.883	0.874	0.91	0.883	0.892	0.894	0.908	0.881	0.888	0.894	0.916	0.885
kr-vs-kp	0.906	0.917	0.92	0.883	0.921	0.924	0.924	0.915	0.923	0.921	0.934	0.91
spambase	0.868	0.871	0.89	0.86	0.874	0.874	0.89	0.871	0.874	0.877	0.899	0.868

Table 5: Accuracy of each delegation method over several ensemble sizes and number of gurus. Accuracy tends to be higher in *max* column but not significantly so.

Full: A standard ensemble with all voters equally weighted. This corresponds to a setting with no delegations.

Best: An exhaustive search of all possible delegations from the initial set of voters is performed and the delegations that lead to the highest accuracy are used, regardless of the number of gurus.

Diverse: Each pair of voters is scored according to the q-statistic diversity metric. In the most correlated pair of voters, the less accurate voter delegates to the more accurate. If that delegation would create a cycle the pair is skipped and the next most correlated pair is considered. This repeats until the number of gurus is equal to the predetermined amount, n^{final}.

Max: The n^{final} most accurate voters serve as gurus and all other voters delegate to them dividing their weight as uniformly as possible.

Random: Voters are removed from the ensemble at random until the number of gurus is equal to n^{final}. This is meant to serve as a control and indicate whether Diverse and Max strategies are useful delegation methods.

Nash: In arbitrary order each voter considers each delegation they can make (including to themselves as a guru) and uses the one they can make that will most improve group accuracy. This is repeated until no voter wishes to change their delegation regardless of the number of gurus.

5.3 Results

We present the results of three experiments, each designed to evaluate our system from a different perspective. First, we compare the **Best** delegation mechanism to other mechanisms to determine whether, in the most optimistic case, it is common for delegation to lead to a significant accuracy improvement. Our second experiment considers larger ensembles with varying size and number of delegations to determine whether they become more beneficial with more possible delegations. Finally, we explore the case when voters aim to greedily maximize group accuracy until reaching an equilibrium to better understand how often strictly beneficial delegations exist. Together, these experiments analyze the optimal outcomes from delegation, present a range of more efficiently achievable results, and show how often self-interested agents might be likely to delegate.

5.3.1 Optimal Delegations. We first show the maximum potential benefit of delegations. Table 4 shows results of experiments when comparing across each delegation mechanism. Results are averaged for each dataset over ensembles with initial sizes of 5, 7, and 9 and each of 3 or 5 gurus. Due to the large number of possible delegations it is impossible to examine all delegations in ensembles of larger sizes. When best delegation results are compared with the full ensemble there is no significant difference in their scores on each test fold using a t-test with $p < 0.05$. This shows a lack of meaningful evidence for the benefit of liquid democracy in a real-world situation.

5.3.2 Full Ensembles. Considering larger ensembles paints a similar picture. Table 5 compares results with up to 49 voters and 25 gurus. As n increases there is often a small increase in the accuracy from delegation methods but it fails to become significant at any size. The accuracy of all ensemble types increases with size, however, the differences between each delegation method stay relatively constant. While the max delegation strategy often leads to higher accuracy than others, the majority of these results, as before, are not significantly different from the full ensemble accuracy on a t-test with $p < 0.05$.

5.3.3 Delegation to Equilibrium. We also consider what outcomes arise when voters are able to choose whether or not to delegate until reaching equilibrium. Section 4 describes this process for single-issue voters; extended to a classifier setting this means that after each voter is trained they calculate what ensemble accuracy would be for every possible single delegation they may make. If any delegations lead to a strict improvement in accuracy they use the one giving the most improvement. This is repeated for each voter, sequentially, until no voter wishes to change their delegation.

Table 6 shows that delegation is very infrequently required for ensembles to reach equilibrium. As the number of voters grows, it becomes increasingly likely that the initial ensemble will not strictly increase its accuracy from any single delegation. The average boost in accuracy over all experiments in Table 6 is approximately 0.007, showing further lack of meaningful benefit from delegation.

6 DISCUSSION

This paper has shown that liquid democracy can improve the accuracy of a group at identifying ground truth in a basic theoretical setting. However, when we extend our model to a more practical classification task this improvement fails to materialize. Specifically, we have seen that under strong assumptions of agent knowledge, delegation can always reach an equlibrium with weakly higher group accuracy in a traditional ground-truth voting setting and

dataset	9	19	29	39	49
breast-cancer-w	0.44	0.34	0.26	0.06	0.18
credit-approval	0.12	0	0.02	0	0.02
heart	0.52	0.24	0.24	0.2	0.1
ionosphere	0.34	0.3	0.28	0.1	0.14
kr-vs-kp	0.32	0.16	0.16	0.12	0.14
spambase	1.24	0.82	0.46	0.54	0.36

Table 6: Average number of delegations required for voters to reach a Nash equilibrium for increasing numbers of voters.

will almost always allow for weak improvement of accuracy when extended to a machine learning setting. Experimentally we have seen that even optimal delegations do not significantly improve accuracy in small ensembles. In more realistically sized ensembles, no efficiently computable delegation strategy signficantly improves accuracy. We also considered the (perhaps more realistic) situation where voters only care about the direct effect of their delegation. This led to our highly informed voters simply not delegating as delegation would not lead to immediate benefit.

While our model may not be directly applicable to real world voting settings involving humans, our results do raise questions about the practical benefit of delegations in such elections. If voters do not see a clear benefit to themselves from delegating why should they do it? A common answer is that voting directly can be costly, but would simply making direct voting a more practical option not provide a more ethical outcome?

6.1 Future Work

Direct followup to this work can focus on providing more explicit theoretical bounds on the expected gain in accuracy from delegations. General models of liquid democracy with unrestricted delegations have received little such analysis, as providing any clear results is difficult. Another useful approach would be to provide stronger axiomatic bounds on various forms of liquid democracy, early work on this has been done in papers such as [12, 22].

More generally, however, more thought must be given to where liquid democracy can be applied. Many papers present competing solutions to numerous problems caused by delegation such as the rise of dictators or delegation cycles, often relying on centralized mechanisms to determine delegations. Such mechanisms may not be realistic – will any human voter delegate when they don't know where there delegation will end up? If not, are there applications (such as machine learning) where delegation can be shown to provide a benefit? Future research must address questions such as these and should provide clear empirical evidence for the benefit of liquid democracy.

REFERENCES

[1] Muesluem Atas, Thi Ngoc Trang Tran, Ralph Samer, Alexander Felfernig, Martin Stettinger, and Davide Fucci. 2018. Liquid Democracy in Group-based Configuration.. In *ConfWS*. 93–98.

[2] Jan Behrens, Axel Kistner, Andreas Nitsche, and Björn Swierczek. 2014. *The principles of LiquidFeedback*. Interaktive Demokratie e. V. Berlin.

[3] Daan Bloembergen, Davide Grossi, and Martin Lackner. 2019. On rational delegations in liquid democracy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1796–1803.

[4] Christian Blum and Christina Isabel Zuber. 2016. Liquid democracy: Potentials, problems, and perspectives. *Journal of Political Philosophy* 24, 2 (2016), 162–182.

[5] Paolo Boldi, Francesco Bonchi, Carlos Castillo, and Sebastiano Vigna. 2011. Viscous democracy for social networks. *Commun. ACM* 54, 6 (2011), 129–137.

[6] Paolo Boldi, Corrado Monti, Massimo Santini, and Sebastiano Vigna. 2015. Liquid FM: recommending music through viscous democracy. *arXiv preprint arXiv:1503.08604* (2015).

[7] Markus Brill. 2018. Interactive democracy. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 1183–1187.

[8] Markus Brill and Nimrod Talmon. 2018. Pairwise Liquid Democracy.. In *IJCAI*, Vol. 18. 137–143.

[9] Gavin Brown, Jeremy Wyatt, Rachel Harris, and Xin Yao. 2005. Diversity creation methods: a survey and categorisation. *Information Fusion* 6, 1 (2005), 5–20.

[10] Ioannis Caragiannis and Evi Micha. 2019. A Contribution to the Critique of Liquid Democracy.. In *IJCAI*. 116–122.

[11] Zoé Christoff and Davide Grossi. 2017. Binary voting with delegable proxy: An analysis of liquid democracy. *arXiv preprint arXiv:1707.08741* (2017).

[12] Rachael Colley, Umberto Grandi, and Arianna Novaro. 2021. Smart Voting. In *Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI 2020)*. International Joint Conferences on Artifical Intelligence (IJCAI), 1734–1740.

[13] Cristina Cornelio, Michele Donini, Andrea Loreggia, Maria Silvia Pini, and Francesca Rossi. 2019. Voting with Random Classifiers (VORACE). *arXiv preprint arXiv:1909.08996* (2019).

[14] Charkes L. Dodgson. 1884. *The principles of parliamentary representation*. Harrison and Sons.

[15] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

[16] Paul Gölz, Anson Kahng, Simon Mackenzie, and Ariel D Procaccia. 2018. The fluid mechanics of liquid democracy. In *International Conference on Web and Internet Economics*. Springer, 188–202.

[17] James Green-Armytage. 2015. Direct voting and proxy voting. *Constitutional Political Economy* 26, 2 (2015), 190–220.

[18] Steve Hardt and Lia CR Lopes. 2015. Google votes: A liquid democracy experiment on a corporate social network. (2015).

[19] Yili Hong. 2013. On computing the distribution function for the Poisson binomial distribution. *Computational Statistics & Data Analysis* 59 (2013), 41–51.

[20] Anson Kahng, Simon Mackenzie, and Ariel D Procaccia. 2018. Liquid democracy: An algorithmic perspective. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[21] Christoph Carl Kling, Jérôme Kunegis, Heinrich Hartmann, Markus Strohmaier, and Steffen Staab. 2015. Voting behaviour and power in online democracy: A study of LiquidFeedback in Germany's Pirate Party. *arXiv preprint arXiv:1503.07723* (2015).

[22] Grammateia Kotsialou and Luke Riley. 2018. Incentivising Participation in Liquid Democracy with Breadth-First Delegation. *arXiv preprint arXiv:1811.03710* (2018).

[23] Ludmila I Kuncheva and Christopher J Whitaker. 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning* 51, 2 (2003), 181–207.

[24] James C Miller. 1969. A program for direct and proxy voting in the legislative process. *Public choice* 7, 1 (1969), 107–113.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[26] Yuzhe Zhang and Davide Grossi. 2020. Power in Liquid Democracy. *arXiv preprint arXiv:2010.07070* (2020).