Liquid Democracy in Practice: An Empirical Analysis of its Epistemic Performance

Manon Revel
mrevel@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Daniel Halpern
dhalpern@g.harvard.edu
Harvard University
Cambridge, Massachusetts, USA

Adam Berinsky
berinsky@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Ali Jadbabaie
jadbabai@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

ABSTRACT

Liquid democracy is a voting paradigm that allows voters to either vote themselves or delegate (eventually transitively) their vote to a peer. Votes are counted through a weighted majority, where a delegate's weight equals the number of individuals they represent post-delegation. Liquid democracy promises to enhance collective decisions through a process deemed both legitimate (delegates are chosen endogenously by all) and accurate (experts tend to receive more delegations). Such assertions rely on both delegations improving the group's expertise post-delegation and no delegate amassing too much power. To investigate liquid democracy on binary issues for which there is a ground truth, Halpern et al. [16] modeled delegation behavior stochastically and identified sufficient conditions such that liquid democracy performs better than direct democracy. Herein, we investigate whether these conditions are met empirically. Through six experiments with a total of $N = 101$ participants from 14 countries, we test the performance of liquid democracy by asking voters to either vote or delegate on five *tasks* (group of questions from a unique theme). Regardless of their delegation choices, we collect voters' answers to all questions and compare the liquid vote with its counterfactual, the direct vote. We observe that higher-expertise voters are statistically less likely to delegate than lower-expertise ones. Further, the average expertise of voters who delegate is lower than the expertise of those receiving delegations. These findings are aligned with Halpern et al.'s requirement and empirically suggest that delegation behaviors meet the conditions for positive theoretical guarantees.

KEYWORDS

liquid democracy, Wisdom of Crowds, epistemic democracy, representation

1 INTRODUCTION

Liquid democracy is a voting paradigm that allows voters to either cast a vote or nominate a delegate to decide on their behalf. Delegations are transitive so that if A delegates to B, B delegates to C, and C votes herself, C effectively casts a vote on behalf of all three. The final decision is made through a weighted majority where a voter's weight equals the number of delegations she received; this is illustrated in Figure 1.

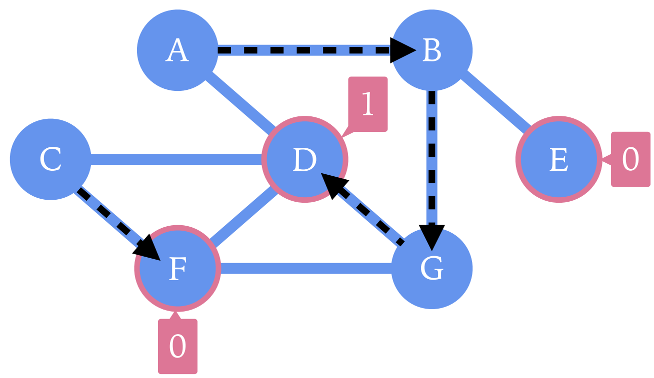

Figure 1: Liquid vote between propositions 0 and 1.

The figure represents the output of a liquid vote on two propositions (0 and 1) among $N = 7$ voters. The voters are connected through an underlying social structure illustrated by the blue lines. The dotted black arrows represent delegations: voter A delegates to voter B. The voters who delegate (voters $A, B, C,$ and G) are called delegators. The voters who vote directly, that is, do not delegate and take part in the final vote (voters $D, E,$ and F, circled in pink) are called delegates. In liquid democracy, votes are counted through a weighted majority where each voter $i \in [N]$ has weight w_i depending on their delegation behavior. Each delegate's weight equals the number of individuals they represent either directly or transitively (here $w_D = 4$, $w_E = 1$ and $w_F = 2$). Indeed, delegate D represents herself along with voters G, B and A. Delegate E, on the other hand, solely represents herself. The weight of the delegators in the final decision is equal to zero. Finally, the pink boxes display the policy for which each delegate votes. The decision is made among the delegates, each of their vote weighted by the number of delegations they received. Proposition 0 hence gathers $w_E + w_F = 3$ votes and proposition B $w_D = 4$ votes. In summary, this liquid assembly chose proposition 1.

Liquid democracy has been said to combine the best aspects of direct voting (where all voters cast a vote) and representative democracy (where voters elect representatives to vote on their behalf) [4]. Moreover, it is currently being proposed as an alternative to existing voting practices to elect per-issue bodies of experts (or congress-members) [40]. Evaluating such proposals is beyond the scope of this paper; instead, we investigate the empirical performance of liquid democracy on closed questions, i.e., those with a correct answer. While such results cannot alone be used to advocate for or against liquid democracy, they would test a key assumption at the heart of this voting paradigm: local delegations will find experts in the electorate and lead to better decisions. We will focus on the *epistemic* setting, where voters decide on a binary issue for which there is a ground truth, and evaluate the epistemic dimension of decision-making investigating the performance of various rules in identifying the correct answer to given problems.[1]

1.1 Collective Intelligence

Researchers on the epistemic dimension of collective decision-making have documented for over two centuries the power of *collective intelligence* that emerges when a group, through its collective agency, is wiser than any of its individual members. These results have theoretical underpinning [9] as formalized mathematically by the Marquis de Condorcet in 1785 (in what is known as the Condorcet Jury Theorem) and have been supported by considerable empirical evidence [12, 37], philosophical argument [24], and use in business applications, for instance, in prediction markets [1] and crowdsourcing [10, 41].

In the simple case of N voters facing a binary decision (think of the question "Will Emanuel Macron or Marine Le Pen win the French presidential election?") where 0 represents the wrong answer and 1 the correct answer, a priori unknown, this phenomenon can be modeled as follows. We use the notation $[k] = \{1, \ldots, k\}$. Each voter $i \in [N]$ has a *competence* or *expertise* level $p_i \in [0, 1]$ (we will use these terms interchangeably). This p_i represents the probability the voter votes correctly, hence, their vote is a sample $X_i \sim \mathrm{Ber}(p_i)$. In general, we will assume that these votes are mutually independent. Further, we will assume that the p_is are themselves drawn i.i.d. from some distribution $\mathcal{D}$. Note that the proportion of correct votes will approach $\mathbb{E}[\mathcal{D}]$ as N increases (simply from the Law of Large Numbers). Hence, if the average expertise of a group member $\mathbb{E}[\mathcal{D}]$ is strictly greater than 0.5, the probability that at least half of the voters are correct converges to 1 as N increases. In other words, for N large enough, even when no individual citizen is perfectly accurate, the group almost certainly converges on the correct answer. [2]

Of course, this result is flipped should the average expertise of a group member $\mathbb{E}[\mathcal{D}]$ be strictly smaller than 0.5. Empirically, groups are also known to fall into this sub-optimal regime, creating what James Madison referred to as the *confusion of the multitude* [29].[3]

The performance of this collective agency is therefore known to depend on the inner characteristics of the groups. In lay terms, for binary decision making, the premise of collective intelligence is that the average group member is better at voting than a random fair coin. Importantly, liquid democrats suggest displacing the necessary condition of collective intelligence from "knowing about an issue" to "knowing who knows about this issue." I might not know whether proposition 0 or proposition 1 is better suited to curb climate change, but I most likely know who knows more than me. Liquid democracy could leverage collective intelligence to identify the knowledgeable agents and increase the likelihood of being collectively correct. Herein, we identify whether such a phenomenon (where voters identify more competent others through delegation) happens in practice.

Now, even if liquid democracy were to increase the expertise of the average group member endogenously through delegations, it may also lead to excessive concentration of power where, in the extreme case, a unique delegate receives all the delegations. Beyond the philosophical concerns, the Condorcet Jury Theorem suggests that such a situation would be mathematically sub-optimal. Along these lines, Kahng et al. [20] proved that, under a certain class of delegation behaviors, it is always possible to construct pathological network typologies such that a few agents amass too much power for liquid democracy to outperform a majority vote. Hence, besides testing whether voters identify more competent others, we will also comment on whether we observe concentration of power in our studies.

1.2 Related Work

Liquid Democracy. In Halpern et al. [16], the authors precisely studied this trade-off identifying sufficient conditions on the maximum number of delegations one may receive and the average increase in expertise post-delegation for liquid democracy to outperform direct democracy.

They further identify types of delegation behaviors that lead to liquid assemblies whose characteristics respect the trade-off mentioned above. They model delegation as dependent on voters' relative expertise. Concretely, they consider a function $q : [0, 1] \to [0, 1]$ that maps expertise to probability of delegation so that voter i with associated competence p_i votes with probability $q(p_i)$. Next, if voter i delegates, she samples a peer j to delegate to with probability proportional to a value $\varphi(p_i, p_j)$ where $\varphi : [0, 1]^2 \to [0, 1]$ depends on both delegator i and potential delegate j's expertise and outputs the probability that this neighbor is chosen. The authors show that the following three classes of delegation behaviors are sufficient for liquid democracy to weakly outperform direct democracy:

- **Upward delegation:** Voters delegate with a fixed probability p independent of their expertise but only delegate to more competent peers. In short: for all $i \in [n]$, $q(p_i) = p$ and for all $(i, j) \in [n]^2$, $\varphi(p_i, p_j) = 1_{\{p_j > p_i\}}$.
- **Confidence-based delegation:** Voters' propensity to delegate decreases with their expertise, and they choose someone randomly when they delegate. In short: $q(x)$ is a decreasing function and for all $(i, j) \in [n]^2$, $\varphi(p_i, p_j) = 1$.
- **General Delegation:** Voters delegate with a fixed probability p independent of their expertise, but they put higher weight on more competent peers. In short: for all $i \in [n]$, $q(p_i) = p$ or $q(x)$ is a decreasing function and $\varphi(x, y)$ increases in its second coordinate.

[1] We will solely focus in this paper on issues that have a unique correct answer. In Section 4, we will discuss the extent to which such insights extend beyond this to decisions that are not only fact-based but moral-based.

[2] Note that this result still holds when comparing the group to the highest order statistic of a fixed distribution (not changing with N) that has 1 in its support, as the probability that a sum of Bernouillis converges grows exponentially fast while the probability that the highest order statistics does grows as $1/N$ [34].

[3] The death of the philosopher Socrates is often taken as an example of collective confusion. Socrates was put on trial for "corrupting the youth" by politicians unhappy with Socrates' effort to teach students to have a critical spirit and sentenced to death by a majority vote (56%) of 501 Athenians [31].

The point of the present paper is to investigate the validity of the delegation behaviors identified in [16]. Precisely, we will test whether voters delegate more often when they are less competent and when delegating, whether they tend to choose more competent agents.

While [2, 7, 20] exposed negative results for liquid democracy exhibiting pathological graphs with an intolerable amount of concentration of power and proving hardness results when trying to find the optimal delegation flows, Halpern et al. identify delegation behavior assuming connected social structure such that liquid democracy proves to be a better-performing voting system than direct democracy.

Note that liquid democracy has further been studied through many lenses other than this epistemic one. From a political economy perspective, Green-Armytage [15] studies whether utility-maximizing agents would rationally delegate; Bloembergen et al. [3] and Zhang and Grossi [42] analyze more sophisticated game-theoretic frames to motivate both voters and delegates' rationale in liquid democracy. Christoff and Grossi [8] investigate logically interdependent propositions connecting liquid democracy to a DeGroot model where voters "copy" their neighbor's signal.

Others have proposed various practical solutions to bypass empirical hurdles associated with liquid democracy: Brill and Talmon [5] proposed to let voters nominate multiple delegates in case some abstain and also suggests ways to let a central planner decide who would receive the delegation among the short-list. In a similar vein, Gölz et al. [14] offers to let voters nominate k delegates and rely on a central planner to choose the delegation graph that would minimize concentration of power.

Finally, political philosophers have been studying the normative properties of liquid democracy [4, 39] and proposing it as an alternative to the current legislative processes [25, 40]. Such research often follows [23]'s minority view that representative democracy, if achieved through cogent elections, may be a greater form of democracy than direct democracy.

Wisdom of Crowds in Practice. This paper provides an empirical analysis of liquid democracy's performance, also relying on the empirical literature focused on collective behaviors and the "psychology of crowds" [26].

In his controversial 1895 book, Gustave Le Bon, with the spectrum of the French Revolution in mind, defines different types of crowds and rationalizes their predictable irrationality through the concept of "popular mind." In his *Memoirs of Extraordinary Popular Delusions and the Madness of Crowds*, Charles Mackay references instances where groups' judgements resulted in disastrous outcomes [28]. Yet, examples dating back to the early twentieth century have exhibited a phenomenon called "Wisdom of Crowds" in which the quality of groups' judgements surpasses those of a few experts. For example, Francis Galton famously collected 787 predictions for the weight of an ox and observed that the "median of the guesses—1,207 pounds—was, remarkably, within 1% of the true weight" [12, 36].

Such experiments have been repeated over the years to assess knowledge [37], forecast stock prices [22], identify phishing websites [27, 30], forecast political or social events [6, 19, 21, 32], and predict sporting outcomes [17]. Predictions markets [1] have also promised to deliver more accurate forecasts on forums generating revenues for the prediction of highly uncertain events. Crowdsourcing also was built on similar premises [10, 41].

The most comprehensive empirical investigation of the Wisdom of Crowds, to the best of our knowledge, is by Simoiu et al. [36]; they collected around 500,000 answers from almost 2,000 participants for about 1,000 questions spanning 50 different domains ordered in 5 categories (knowledge, tacit, popular culture, predictions, and spatial reasoning). They found that the crowd does better on individual questions and "considerably better than individuals when performance is computed on a full set of questions within a domain."

While these experiments use simple aggregation metrics (such as the mean or median of a sample), others were tested in an attempt to extract an even wiser substrate from the signals gathered from a group, such as in Prelec's *truth serum* [33].

Wisdom of Crowds has also been found to fail in certain setups, for example, situations "in which emotional, intuitive responses conflict with more rational, deliberative responses" [36]. For instance, Simmons et al. [35] found that voters' biases prevented them to make wise decisions in the sports betting context.

The notion of the Wisdom of Crowds is not, as it may first seem, at odds with the idea of expertise. On the contrary, researchers have identified that often, small crowds of identified *experts* perform better than the large less-informed crowds [6, 13, 18, 38]. Liquid democracy promises to identify such a smaller crowd endogenously.

1.3 Experiment Goals

In what follows, we present a series of six experiments where voters were given the chance to either vote or delegate different tasks. Even in the case that they delegate on a given task, they were still asked the question in the second phase of the experiment. This allows us to do a few things. First and most directly, we can compare the accuracy of voting under liquid and direct democracies. Second, we can use the answers to all questions to get an estimate of voter competencies by simply considering their number of correct answers. From this, we can study how delegation behavior depends on expertise. Our hypothesis is that the behaviors match the sufficient theoretical conditions of Halpern et al. [16]; that is, first, voters are more likely to delegate the less competent they are, and second, on average, more competent voters tend to receive more delegations. Finally, we hope to experimentally exhibit that many voters choose to vote directly and delegations are relatively balanced, preventing accuracy-harming concentration of power. The paper is organized as follows. In Section 2, we describe the survey content (Section 2.2), the participants' characteristics (Section 2.1), and the statistical models used for inference (Section 2.3). Next, in Section 3, we present the results looking at high-level delegation behaviors (Section 3.1) before digging into the relationship between expertise and delegation, both in likelihood of delegating in Section 3.2 and in choice of delegate in Section 3.3. We conclude the section comparing the performance of liquid democracy and direct democracy (Section 3.4). Finally, Section 4 discusses the experiment's limitations and routes for future work.

2 METHODS

In this section, we present the different participating groups, the survey material, and the analysis strategy.

2.1 Experiments

We conducted $E = 6$ experiments between March 21st and April 5th, 2022, with a total of $N = 101$ participants from 14 different countries in Asia, Africa, North and South America, and Europe. Of the participants across all experiments, 29% were native English speakers, 16% were female, 4% non-binary, and 80% were male. The series of experiments were conducted in Morocco, the United States, France and the Netherlands. Each experiment involved one group of 11 to 32 participants that had some social ties. (We will use interchangeably the words “group” and “experiment” as again, each experiment was conducted with a different group.) The details of the groups that participated in the study are displayed below in Table 1.

In each experiment, group members answered a survey with the same 25 questions grouped into 5 tasks. The questions were also identical across experiments, except for the prediction task, which involved predicting sporting events that were to take place in the following week (so the question needed to be changed periodically). Section 2.2 is dedicated to detailing the survey content.

Note that such experiments cannot be run anonymously online since liquid democracy relies on groups having an underlying social structure requiring more involved organization. We partnered with universities, companies, and associations that helped us access such groups. The experiment was submitted to a university Committee on the Use of Humans as Experimental Subjects and is IRB exempted.

2.2 Material

2.2.1 Questions. Each participant was faced with 5 tasks. A task involved answering a series of questions from a given domain. The 5 domains, taken from Simoiu et al. [36], were: knowledge (identify historical landmarks), pop culture (link a music theme to a movie), tacit (recognize English idioms), prediction (predicting sporting events), and spatial reasoning (following the position of a hidden ball under cups). Each participant answered the same series of questions (3 to 7) within each task.

In total, there were 25 questions grouped in the different tasks (see Table 2). The questions were taken from Simoiu et al. [36] that developed a curated list of 100 questions grouped into 50 domains, further stratified into the five categories (knowledge, pop culture, tacit, prediction, and spatial reasoning). To be sure, Simoiu et al. [36] had multiple domains per category. We picked a unique domain per category and created the tasks around that domains.[4] To be consistent with the epistemic setup under study, we converted all categorical questions into binary ones. As an example, for a question from Simoiu et al. [36] of the type “Where is this famous landmark from?” with four options (Italy, Tibet, Greece, or Brazil) to choose from, we selected a random option to reformulate the question as: “Is this famous landmark from Brazil?” Finally, note that the prediction questions from Simoiu et al. [36] were about March Madness (a basketball tournament in the United States for college students playing first division of NCAA), and we re-used those for our first two experiments. Then, for the experiments after March Madness concluded, we constructed questions about other upcoming sporting events. The questions are shown in Table 2.

2.2.2 Survey Flow. For each group, a survey link was provided. The participants were first asked for informed consent and names. They were then given a description of a *task* and were asked whether they wanted to delegate or not. If not, they were taken to all the questions from this directly. On the other hand, if they chose to delegate, they were asked to select the name of their delegate and were then presented with the next task. The prompts for each task are shown here:

- **Knowledge**: You will be shown images of architectural landmarks from around the world, and asked to select the country where the landmark is located.
- **Popular Culture**: You will be provided with short audio files with theme songs from various movies, and asked to select the movie it was featured in.
- **Tacit**: You will be given English idioms, and asked to identify their meaning. An idiom is a group of words that have a meaning not deducible from those of the individual words (e.g., rain cats and dogs, see the light).
- **Prediction**: You will be given US college basketball teams, and asked to predict which round they will make it to in the NCAA Tournament, taking place in March 2022. OR [5] You will be given upcoming soccer games, and asked to predict the games’ outcome. OR You will be given upcoming sporting events (soccer and tennis games), and asked to predict the games’ outcome?
- **Spatial Reasoning**: You will be asked to watch a short video of the Cups and Balls magic trick, and identify the location of the ball at the end of the trick.

After being exposed to all five tasks and asked to vote or delegate, participants were taken to the final stage of the survey, where they were asked to answer “additional questions” that were, in fact, all the questions they had delegated in the first stage. This was done at the end of the experiment, not to prime the participants on the exercise. These signals allowed us to compare the results of liquid democracy with direct democracy, the counterfactual where all voters had voted. Finally, a few optional background questions were asked on the last page. Excerpts from the survey are shown in Figure 6.

Note that we randomized the order in which tasks, questions within each task, and the “True/False” options appeared.

Next, we turn to explaining the ways data were processed and analyzed.

2.3 Analysis Strategy

We begin with notation that we will use throughout the paper. Let $[N]$ be the set of N voters and $[E]$ be the set of E experiments. Each experiment $e \in [E]$ has N_e participants so that $N = \sum_{e \in [E]} N_e$.

[4] We selected a random discrete domain per category and, for each domain, chose a random sample of questions. In a discrete domain, participants chose between a finite set of pre-set answers instead of answering open-ended questions.

[5] Again, different prediction questions were used for different experiments, because predicted outcomes were realized between the running of experiments.

Table 1: Groups Characteristics

Qualitative groups description, sizes and performance under direct and liquid democracy across all tasks

Group ID	Group Description	Group Size	Direct Democracy	Liquid Democracy
1	Research Group: Members connected for 1 to 10+ years	11	0.682	0.713
2	Graduate and Undergraduate Class: Members from various years and programs that do not necessarily know each other outside of class	12	0.726	0.730
3	Graduate Class: Students from various years in a program with small pods and little interconnections among them	32	0.682	0.695
4	Sports Team: Members connected for 3 months to 10+ years	14	0.740	0.748
5	Financial Association: Members with shared interests and common enterprises for 2+ years	17	0.694	0.729
6	Group of employees, students and faculty: Members connected for 1 to 10+ years with various degrees of connection	15	0.678	0.703

Let $\mathcal{T} = \{P, PC, SR, T, K\}$ be the set of five tasks (respectively Prediction, Popular Culture, Spatial Reasoning, Tacit, and Knowledge). For each task $t \in \mathcal{T}$ there are $[R_t]$ questions which prompts can be found in Table 2. For each voter $i \in [N]$ in experiment $e \in [E]$ and for each question $r_t \in [R_t]$ in each task $t \in \mathcal{T}$, we collect:

(i) the direct vote to each question $v^D_{i,e,r_t} \in \{0, 1\}$ (votes are binary)
(ii) the binary signal $\delta_{i,e,t}$ taking on the value 1 if i delegated the task t and 0 otherwise
(iii) the liquid vote $v^L_{i,e,r_t} \in \{0, \cdots, N_e\}$
(iv) the confidence level for each question r_t, denoted by $c_{i,e,r_t} \in \{0, 0.25, 0.5, 0.75, 1\}$

Indeed, recall that all participants answered every question (either in the first experimental phase after choosing "Vote Myself" or in the second experimental phase in the "additional questions" section); this corresponds to v^D_{i,e,r_t}. Then, every participant was asked in the first experimental phase to either vote or delegate *at the task level*, which is a binary behavior captured by $\delta_{i,e,t}$. The liquid vote represents the contribution of voter i to the liquid vote on question r_t, that is i's vote multiplied by i's transitive weight post-delegation w_{i,e,r_t} (see Figure 1 for examples of weights). Hence, $v^L_{i,e,r_t} = v^D_{i,e,r_t} \times w_{i,e,r_t}$ the liquid vote is 0 if voter i delegates, and i's weighted vote otherwise. Finally, we measure voter confidence by asking, after each question r_t, how confident the voters felt with their answer on a 5-point scale (normalized so that 0 is "not confident at all" and 1 is "extremely confident").

To be sure, the quantities of interest are: the votes of each voter i on each question r_t (where v^D_{i,e,r_t} is the direct vote and v^L_{i,e,r_t} is the liquid vote) and the delegation decision voter i took at task t, δ_{i,e,r_t}. Note that the indexes i and e are redundant but we keep track of both so that we can account for the heterogeneity between individuals and groups' votes.

If a voter received delegations but did not participate in the study, the delegations were ignored. In the one instance of a cycle (voter i delegated to voter j who delegated to voter i), the delegations were also ignored; in many real-world implementations, such voters would be notified of the cycle and asked to choose a new delegate or vote directly.

Again, the goal of this paper is to study delegation behaviors as we are testing whether the delegation behaviors match the theoretical conditions in [16]. We explain below how we do so.

2.3.1 *Delegate or Not to Delegate: the role of expertise.*

Fixed-effect Models with Robust Clustered Standard Errors. First, we want to test whether voters that do not delegate are statistically more competent than those who do delegate, which would satisfy Halpern et al.'s strongest requirement on their q function (see above). To do so, we compute voter i's expertise in experiment e for task t by finding the proportion of questions voter i correctly answered. That is $p_{i,e,t}$ is the number of correct answers i found among the R_t questions from task t. More formally, voter i's expertise on task t is $p_{i,e,t} = \frac{\sum_{r_t \in [R_t]} v^D_{i,e,r_t}}{R_t}$. We then run a fixed-effect model with robust clustered standard errors, regressing the voter's expertise against their delegation behavior at the task level:

$$v^D_{i,e,r_t} = \alpha_i + \alpha_{r_t} + \alpha_t + \alpha_e + \delta_{i,e,t}\hat{\beta^q} + \varepsilon_i \tag{1}$$

where α_i, α_t and α_e are fixed-effects for the participants, tasks and experiments respectively. The fixed effects capture the fact that there is unobserved heterogeneity between the different groups. Next, ε_i denotes the standard errors that we cluster at the individual-level and at the experiment-level. The robust clustered-standard error account for the fact that the samples within one cluster are not independent and identically distributed (i.i.d.): the answers given by one same voter are correlated. Finally, $\hat{\beta^q}$ is the estimate for the increase in expertise in those who vote versus those who do not delegate.

We run the same analysis stratified per task:

$$v^D_{i,e,r_t} = \alpha_i + \alpha_{r_t} + \alpha_t + \alpha_e + \sum_{\tau \in \mathcal{T}} \mathbf{1}_{\{t=\tau\}}\delta_{i,e,\tau}\hat{\beta^q_\tau} + \varepsilon \tag{2}$$

The standard errors are unclustered (we have a unique sample per voter per task) and $\hat{\beta_\tau}$ is the heterogeneous increase in expertise in those who vote versus those who do not delegate for category τ.

Table 2: Survey Material

Questions used in the liquid democracy survey. Note that different prediction questions were used for different experiments; this is simply because predicted outcomes were realized between the running of experiments. The questions in Knowledge, Popular Culture, and Spatial Reasoning relied on audio-visual documents that we can share upon request.

Category	Prompt	Answer
Knowledge	This landmark is located in Italy.	False
	This landmark is located in Turkey.	True
	This landmark is located in Myanmar.	False
	This landmark is located in France.	False
	This landmark is located in Brazil.	False
Popular Culture	This music was featured as a theme song in the movie The Hobbit.	False
	This music was featured as a theme song in the movie The Empire of Sun.	False
	This music was featured as a theme song in the movie Gravity.	True
	This music was featured as a theme song in the movie Goodfellas.	False
	This music was featured as a theme song in the movie The Pianist.	False
	This music was featured as a theme song in the movie A Passage through India.	False
	This music was featured as a theme song in the movie The Schindler's List.	True
Tacit	"A man of straw" means "A very active person".	False
	"To drive home" means "To emphasize".	True
	"To smell a rat" means "To suspect foul dealings".	True
	"To end in smoke" means "To excite great applause".	False
	"To catch a tartar" means "To deal with a person who is more than one's match".	False
Prediction for Experiments 1-2	The US college basketball team West Virginia Mountaineers will make it to the Elite Eight in the 2022 NCAA Tournament.	False
	The US college basketball team Michigan State Spartans will make it to the First Round in the 2022 NCAA Tournament.	True
	The US college basketball team Syracuse Orange will win the 2022 NCAA Tournament.	False
	The US college basketball team Purdue Boilermakers will make it to the 2nd round in the 2022 NCAA Tournament.	True
	The US college basketball team Arizona Wildcats will make it to the Elite Eight in the 2022 NCAA Tournament.	False
Prediction for Experiments 3-5	Galatasaray SK will beat FC Barcelona during the Europa League game on March 17th.	False
	Olympic de Marseille and OGC Nice will tie during the French League game on March 20th.	False
	VFL Wolfsburg will beat Bayer 04 Leverkusen during the German League game on March 20th.	False
	Salernitana will lose against Juventus during the Italian League game on March 20th.	True
	FC Barcelona and Real Madrid CF will tie during the Spanish League game on March 20th.	False
Prediction for Experiment 6	Eintracht Frankfurt will beat FC Barcelona during the Europa League game on April 7th.	False
	Olympic de Lyon and West Ham United will tie during the Europa League game on April 7th.	True
	Brazil will lose to Spain during the Women's International Friendly game on April 7th.	False
	Neither Rafael Nadal nor Novak Djokovic will qualify for the ATP Masters 1000 Monte Carlo Final on April 17th.	NA
	Stefanos Tsitsipas will win the ATP Masters 1000 Monte Carlo Tournament on April 17th.	NA
Spatial Reasoning	The object is located in the middle cup at the end of the trick.	False
	The object is located in the middle cup at the end of the trick.	False
	The object is located in the right cup at the end of the trick.	True

We repeat the same analysis regressing the voters' average confidence against the binary delegation variable. We denote voter i's expertise on task t by $c_{i,e,t} = \frac{\sum_{r_t \in [R_t]} c_{i,e,r_t}}{R_t}$, where c_{i,e,r_t} is the self-reported confidence disclosed by voter i for question r_t (see above).

Through this model, we test whether the expertise of those who vote directly is significantly greater than those who delegate, and we test whether these results are task-dependent. We repeat the analysis with the confidence levels.

Aggregate behaviors. Finally, we compute voter i's propensity to delegate q_i as the percentage of times she delegates across the five tasks $q_i = \frac{\sum_{t \in \mathcal{T}} \delta_{i,e,t}}{5}$ and voter i's average expertise across tasks $p_i = \frac{\sum_{t \in \mathcal{T}} p_{i,e,t}}{5}$. We regress the later against the former and repeat the same analysis with the average confidence level $c_i = \frac{\sum_{t \in \mathcal{T}} c_{i,e,t}}{5}$. Note that this is not exactly Halpern et al.'s q function as theirs is defined at the task level and ours is an aggregate at the individual's level. We hope this would provide general trends on *how* knowledge (respectively confidence), loosely captured by p_i (respectively c_i), are predictive of aggregate delegation behaviors.

2.3.2 *Whom to delegate to? The role of expertise revisited.*

Fixed-effect Models with Robust Clustered Standard Errors. We then investigate whether voters who delegate choose neighbors that are more competent or less competent. First, we check whether the expertise of those who delegated is worst than the expertise of those to whom they delegate (which would satisfy [16]'s strongest requirement on their φ function (see above)). We only consider the participants that are involved in a direct delegation (either the *delegator* x that delegates, or the *delegate* y that receives the delegation). To do so, we run a fixed-effect model with robust clustered standard errors as in the previous section. We regress the voter's expertise against a binary variable $\kappa_{i,e,t}$ indicating whether they are delegators or delegates.

$$v^D_{i,e,r_t} = \alpha_i + \alpha_{r_t} + \alpha_t + \alpha_e + \kappa_{i,e,t}\hat{\beta^\varphi} + \varepsilon_i \tag{3}$$

where α_i, α_t and α_e are fixed-effects for the participants, and ε_i denotes the standard errors that we cluster at the individual-level and at the experiment-level. The value $\hat{\beta^\varphi}$ is the estimate for the difference between the delegate's average expertise versus the delegator's average expertise.

We run the same analysis stratified per task:

$$v^D_{i,e,r_t} = \alpha_i + \alpha_{r_t} + \alpha_t + \alpha_e + \sum_{\tau \in \mathcal{T}} \mathbf{1}_{\{t=\tau\}}\kappa_{i,e,\tau}\hat{\beta^\varphi_\tau} + \varepsilon \tag{4}$$

Delegate's expertise. Next, for each task t, we compare the expertise of delegator A $p_{A,t}$ with that of delegate B, $p_{B,t}$, to whom A delegated. We plot the delegator's expertise against the delegate's expertise and observe a gradient from less to more competent voters as Halpern et al. hoped it would be.

We repeat the same analysis with the confidence levels $c_{x,t}$ to compare both measurements' power in predicting the delegation behavior.

2.3.3 *Liquid vs Direct Democracy.*

We finally compare qualitatively the results of liquid democracy versus direct democracy. (Each experiment has only a few people so the results are underpowered.) For the group members of a given experiment $e \in [6]$ with N_e participants and a question r_t, we estimate the performance of direct democracy for the question and that group through $d_{e,r_t} = \frac{\sum_{i \in N_e} v^D_{i,e,r_t}}{N_e}$ (the proportion of correct answers across the group) and that of liquid democracy through $l_{e,r_t} = \frac{\sum_{i \in N_e} v^L_{i,e,r_t}}{N_e} = \frac{\sum_{i \in N_e} v^D_{i,e,r_t} \times w_{i,e,t}}{N_e}$ (the weighted proportion of correct answers across the group). We further compute estimates of liquid and direct democracies per experiment across tasks ($d_e = \frac{\sum_{t \in \mathcal{T}} d_{e,r_t}}{5}$ and $l_e = \frac{\sum_{t \in \mathcal{T}} l_{e,r_t}}{5}$ respectively).

3 RESULTS

We will first present a few delegation graphs for each task and experiment before presenting statistical analyses of the delegation behaviors. After this, we analyze the relative performance of liquid and direct democracy in all settings.

3.1 Delegation Graphs

We collected 505 delegation data points, one per participant per task. Of those, 28% were delegations (like delegators A, B, C, G in Figure 1) and 57% are direct voters that did not receive any delegation besides their own (like delegate E in Figure 1). Among the delegates, 21% received only one delegation besides their own (hence had weight 2 in the decision, like delegate F in Figure 1), 11% received two delegations besides their own and just about 1% received five or more delegations besides their own, showing little sign of concentration of power.

Next, we look at the delegation graphs across different tasks and experiments. Recall that delegations happened at the task-level so we represent delegation behaviors per task. For each task t and experiment e, we show in Figure 2 the delegation graphs with all N_e participants in experiments e, represented by nodes labeled by their expertise $p_{i,e,t} = \sum_{i \in [N_e]} v^D_{i,e,t}/R_t$ that is the average number of correct answers given for that task.

The top left plot shows an example of a successful delegation chain to the right, where an expert with $p_{i,6,K} = 1$ was identified by six other voters either directly or transitively through a local expert j with $p_{j,6,K} = 0.8$. On the right, a smaller chain pictures two voters delegating to a more competent expert, who in turn delegates to a non-expert.

Over the course of the six experiments based on five tasks each, we observed only two delegation cycles of size two (where A delegates to B, who delegates to A), both in Experiment 3 with $N_3 = 32$.

The purpose of the next sections is to study these delegation behaviors and understand if statistically significant phenomena emerge in terms of propensity to delegate based on the expertise (Section 3.2) and propensity to identify delegates with high expertise (Section 3.3).

3.2 To Delegate or Not to Delegate? The Role of Expertise.

In this section, we show that those who delegate are statistically less competent than those who do not delegate, as a result of the regression analysis presented in Equation (1), confirming [16]'s stronger hypothesis on the probability to delegate. These results are driven by the delegation behaviors for the tasks tacit and prediction, according to the heterogeneous regression governed by Equation (2). Unsurprisingly, we find that those who do not delegate are more confident in those tasks than those who do not.

Further, when looking at the aggregate delegation behavior across tasks, we observe a negative correlation between delegation

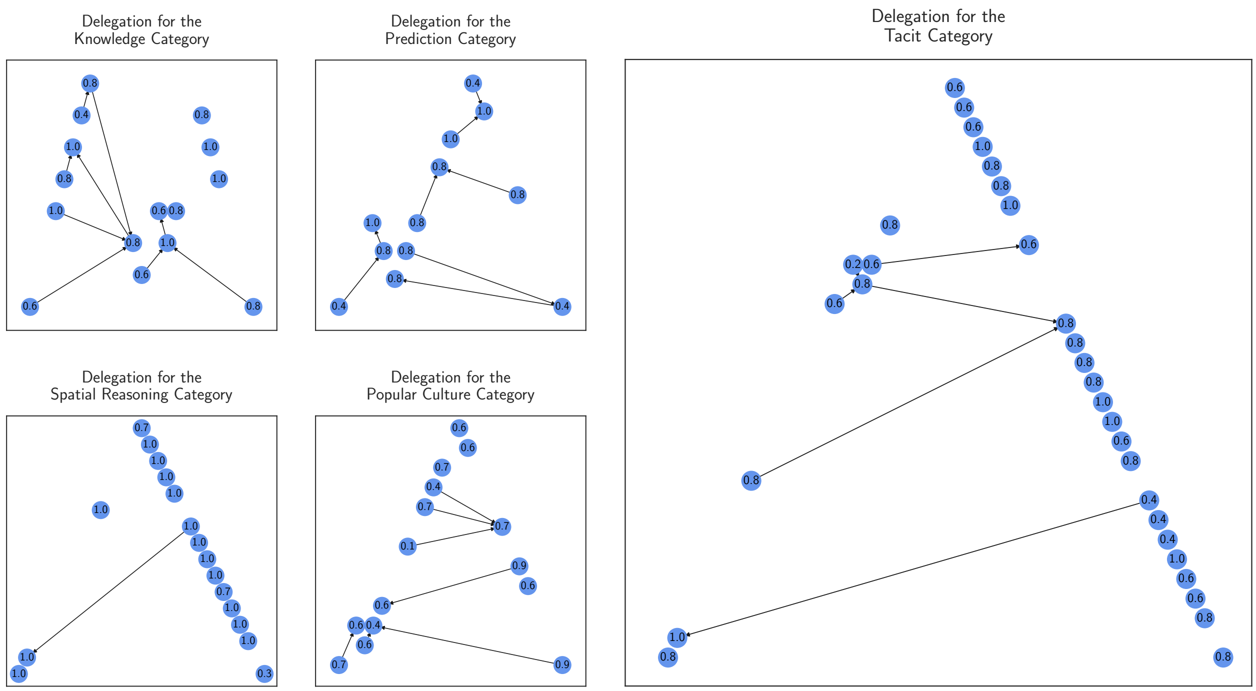

Figure 2: Delegation Graphs for the Different Categories

Each graph represents the performance of a group on a given category. Recall that delegations happen at the task-level, so that one computes the average expertise across tasks $p_{i,e,t}$. Each node represent a voter i in the experiment e and the numbers within the node indicate their expertise $p_{i,e,t}$ on task t. An out-arrow from A to B indicates that A delegated to B. From left to right, top to bottom: Knowledge in Experiment 6*, Prediction in Experiment* 2*, Tacit in Experiment* 3*, Spatial Reasoning in Experiment* 5 *and Popular Category in in Experiment* 6.

rates and expertise, as expected. These results are non-significant. Instead, the delegate rates are significantly negatively correlated with the average confidence.

3.2.1 Significance Tests. Table 3 presents the results of the regression of the average expertise against the binary delegation decision. The estimates displayed are the $\beta^q = 0.078$ (under the Overall Model) and the β_t^qs (under the various Tasks Models) and are the estimates for the difference between the expertise of those who do not delegate versus the expertise of those who do. At the aggregate level, we see that the expertise is significantly lower among those who delegate. This result is driven by the behavior for the Tacit and Prediction tasks.

We compare these results with those regressing the average confidence for each task collected post-delegation for each question and see that those who did not delegate are more confident than those who did across all tasks. These are significant at 95% for the Prediction and Tacit categories too. Note that we expect confidence to be correlated with delegation behaviors.

3.2.2 Aggregate Behavior. Finally, we show the average trend across tasks, testing whether average knowledge and confidence across tasks are negatively correlated with delegation rates. As shown on Figure 3, both average expertise and confidence are negatively correlated with delegation rates, and the regression coefficient is solely significant for the average confidence. Importantly, these quantities do not model the q function we study, as the latter should be defined at the task level. It also shows that, as expected, delegates are more confident than delegators.

Following up on the surveys, we ran interviews and gathered interesting feedback: some voters claimed they never delegated because they made a point to participate directly to be part of the final decision, regardless of their expertise. This speaks to the fact that, while it seems that those who do not delegate are more performing, some voters are driven by other motives when deciding on delegation. While we do not know whether such behavior correlates otherwise with expertise, it indicates delegation behaviors also depend on other procedural factors.

3.3 Whom to delegate to? The role of expertise.

In this section, we only focus on the voters involved in a direct delegation chain of the form "A delegates to B." Recall that we dub A a delegator and B a delegate. Importantly, note that in the previous section, we also accounted for the many voters that voted directly and did not receive any delegation, which are not of interest here.

Table 3: Results on the Relation between Delegation Behaviors and Average Expertise or Confidence

The summary of the average and heterogeneous effects, estimated by the models in Equations (1) and (2). Rows correspond to different contrasts: Average Expertise and Average Confidence. We indicate whether fixed effects were used (t stands for fixed-effects at the task level, e at the experiment level and i at the individual level). We further indicate whether robust clustered standard errors were used to account for correlation within individuals i's answers.

	Overall Model	Tasks Models				
		(Tasks)				
		(Knowledge)	(Prediction)	(Popular Culture)	(Spatial Reasoning)	(Tacit)
Average Expertise $p_{i,e,t}$	0.078^{***}	0.081	0.14^{***}	-0.039	-0.0055	0.19^{***}
	(0.023)	(0.052)	(0.042)	(0.045)	(0.055)	(0.049)
Average Confidence $c_{i,e,t}$	0.17^{***}	0.099^{*}	0.31^{***}	0.099^{*}	0.032	0.21^{***}
	(0.021)	(0.049)	(0.040)	(0.043)	(0.053)	(0.047)
Tasks interaction	No	Yes	Yes	Yes	Yes	Yes
Fixed Effects	t, e, i	t, e	t, e	t, e	t, e	t, e
Cluster-Standard Errors	i	NA	NA	NA	NA	NA
Note:						*p<0.1; **p<0.05; ***p<0.01

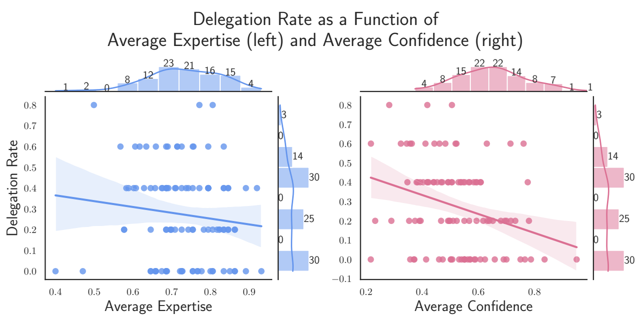

Figure 3: Delegation rate as a function of the average expertise

The left plot represents $q_i = \sum_{t\in\mathcal{T}} \delta_{i,e,t}/5$ the delegation rate as a function of $p_i \sum_{t\in\mathcal{T}} p_{i,e,t}/5$ the average expertise across all tasks for each participant $i \in [N]$. Note that the survey contained five tasks so the delegation rates are stratified. The blue line represents the regression line (slope = -0.24, std = 0.22, p-value = 0.28). The blue line represents the regression line (slope = -0.24, std = 0.2226, p-value = 0.28). This is a loose of estimation of [16]'s q function that maps the probability of delegating to the voter's expertise, with quantities defined at the individual level instead of at the task level. The right plot represents the regression of the delegation rate against the average confidence $c_i = \sum_{t\in\mathcal{T}} c_{i,e,t}/5$. The pink line represents the regression line (slope= -0.46, std= 0.16, p-value=0.0046).

We find that voters statistically delegate to more competent voters, as a result of the regression analysis presented in Equation (3) where we only considered delegators and delegates, confirming [16]'s stronger hypothesis on the delegation function φ. According to the heterogeneous regression governed by Equation (4), these results carry over to the tasks but are only significant for the Tacit task. The difference in confidence between those who delegate and those who do not is even larger and significant across almost all tasks.

We present two heat maps that visualize the previous results, showing, for each level of expertise (respectively confidence), the amount of delegation that goes to a delegate of various expertise levels (respectively confidence). The color gradient indicates that, as hypothesized by [16], delegator A may choose delegates Bs through $\varphi(p_A, p_B)$ increasing in its second coordinate.

3.3.1 *Significance Tests.* Table 4 presents the results of the regression of the average expertise against the binary indicator of whether one is a delegator or delegate. The estimates displayed are the $\beta^{\varphi} = 0.062$ (under Overall Model) and the β^{φ}_t (under the various Tasks Models) and are the estimates for the difference between the expertise of the delegate vs. the delegators. At the aggregate level, we see that the expertise is significantly lower among the delegators. This result is driven by the behavior for the Tacit task. We compare these results with those regressing the average confidence per task collected post-delegation per question.

3.3.2 *Delegate's expertise.* In the previous sub-section, we compared the average expertise of delegators and delegates. Now, we group delegators with the same expertise $p_{i,e,t}$ and investigate the distribution over their delegate's expertise as shown in Figure 4. We see on the left plot that delegators tend to delegate more to more expert agents, corroborating [16]'s strongest requirement on the delegation behaviors. Moreover, we see this phenomenon carries over to the analysis of confidence.

Table 4: Results on the Relation between Delegators and Delegates' Average Expertise or Confidence

The summary of the average and heterogeneous effects, estimated by the models in Equations (3) and (4). Rows correspond to different contrasts: Average Expertise and Average Confidence. We indicate whether fixed effects were used (t stands for fixed effects at the task level, e at the experiment level and i at the individual level). We further indicate whether robust clustered standard errors were used to account for correlation within individuals i's answers.

	Overall Model	Tasks Models				
		(Tasks)				
		(Knowledge)	(Prediction)	(Popular Culture)	(Spatial Reasoning)	(Tacit)
Average Expertise $p_{i,e,t}$	0.062^{**}	0.07	0.044	0.0092	0.053	0.17^{**}
	(0.023)	(0.063)	(0.042)	(0.050)	(0.064)	(0.057)
Average Confidence $c_{i,e,t}$	0.19^{***}	0.21^{**}	0.25^{***}	0.15^{**}	0.023	0.23^{***}
	(0.022)	(0.063)	(0.040)	(0.052)	(0.065)	(0.059)
Tasks interaction	No	Yes	Yes	Yes	Yes	Yes
Fixed Effects	t, e, i	t, e	t, e	t, e	t, e	t, e
Cluster-Standard Errors	i	NA	NA	NA	NA	NA

Note: *p<0.1; **p<0.05; ***p<0.01

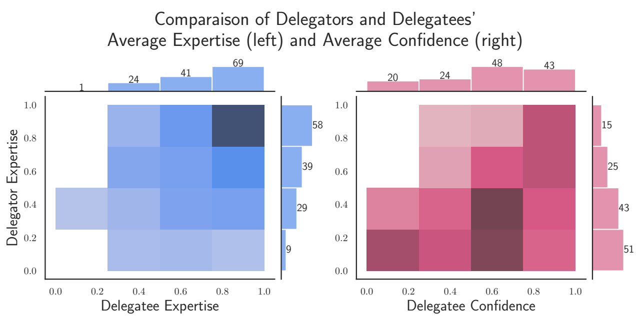

Figure 4: Difference between Delegators and delegates Expertise and Confidence

The left plots represents the number of times a voter with a fixed expertise $p_{i,e,t}$ (bucketed into five equally spaced groups) delegate to a voter with expertise $p_{j,e,t}$ (also bucketed into five equally spaced groups). This is an estimation of a discretization of [16] to check whether voters tend to delegate more to more competent voters. The also embeds the distribution of delegators' (y-axis) and delegates' (x-axis) expertise. The right plot represents the same phenomenon when comparing delegators' and delegates' confidence. It also embeds the distribution of delegators (y-axis) and delegates (x-axis) confidence.

3.4 Liquid or Direct Democracy?

Finally, we compare the performance of direct democracy with that of liquid democracy. For each experiment e and task t, we show the liquid and direct estimates, $l_{e,t}$ and $d_{e,t}$ respectively in Figure 5. Note that none of the sample sizes allow for statistically significant results and the standard errors (not displayed) are rather large. Note that liquid democracy performs rather badly on the task Popular Cateogory. Spatial Reasoning was almost never failed so that the liquid and direct estimates are indistinguishable. The Prediction, Knowledge, and Tacit tasks were usually not as hard as Popular Culture nor as easy as Spatial Reasoning. Liquid democracy consistently improves the decision in Prediction and Tacit tasks, while the pattern is not as clear in the Knowledge tasks. A few observations we made are that (i) liquid democracy tends to help more on questions that are neither too hard nor too easy for the group (the direct democracy score is between 0.6 and 0.8) and (ii) liquid democracy tends to help more on tasks that have a narrower scope (English sentences and soccer games are more alike than movies from any period and landmarks from anywhere in the world).

Table 1 finally shows the difference in estimates d_e and l_e as the performance of liquid and direct democracies per experiment. Liquid democracy is always more efficient than direct democracy, by 0.4% (in Experiment 2, where group members mostly did not know each other) to 3.5% (in Experiment 5, where participants knew each other for 2+ years and share common professional interests).

4 CONCLUSION

This paper focused on the epistemic performance of liquid democracy, where voters decide on a binary issue for which there is a ground truth, and tested the sufficient conditions on the delegation behaviors found by [16]. We find that voters are more likely to delegate when they are less competent and that those who delegate tend to be less competent than those who receive delegations, corroborating the most demanding of Halpern et al.'s conditions.

Interestingly, the delegation behaviors tend to be heterogeneous across tasks and we observe that better-defined tasks (such as Tacit and Prediction) are those on which liquid democracy is most helpful. More research would be needed to test this claim rigorously.

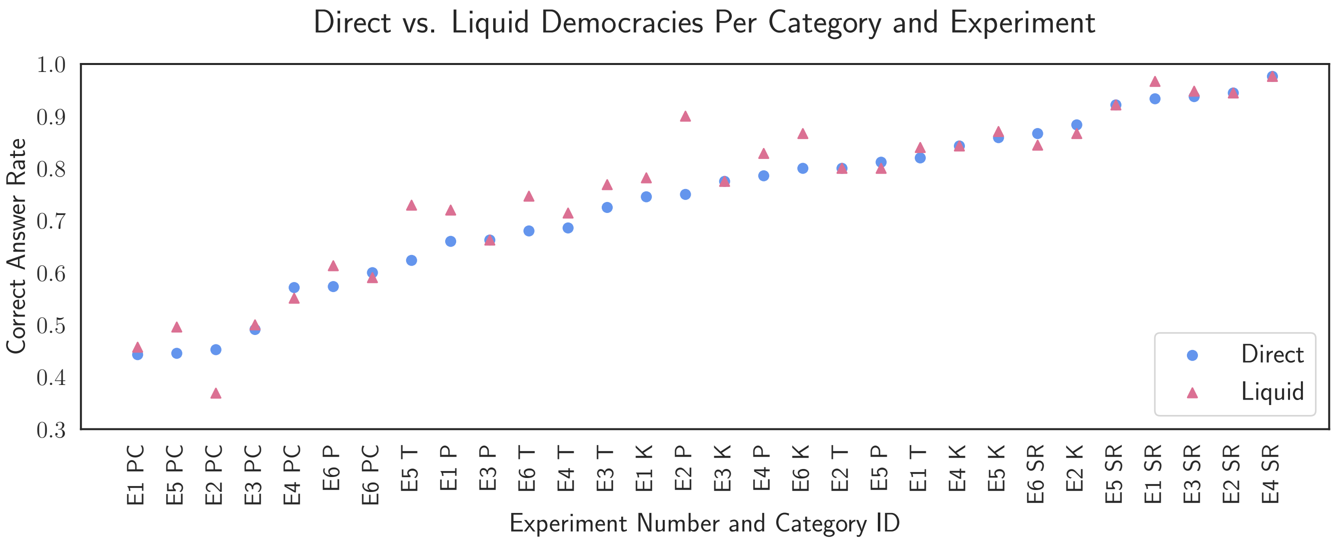

Figure 5: Liquid and Direct Estimates across Tasks and Experiments

For each experiment e and task t, the blue dot (respectively pink triangle) represents the score of the group through direct (respectively liquid) democracy. We observe that liquid democracy tends to do better when direct democracy does neither too bad nor too good and when the scope of the tasks is narrower.

Further, we also used our measure for confidence to analyze whether delegation behaviors could also be driven by self-confidence (under the assumption that charismatic voters known by the group could also attract more delegations because of charisma and not expertise). We find that confidence also increases among those who do not delegate and those who receive delegations. On the one hand, this is intuitive as those who vote directly are likely to do so out of increased confidence (and we see it is also out of higher competence). But, on the other hand, we see that delegations go from statistically less confident voters to more confident voters for certain tasks, in which the expertise is not statistically different between groups. This may suggest that delegations could sometimes be driven by over-confidence instead of expertise. Importantly, note that these results must be taken with a grain of salt as the current experiment was not designed to collect unbiased estimates (confidence was asked *after* voters decided to delegate or not, which might have biased those who did not delegate to claim they were more confident than those who did not). Instead, we believe this is an interesting trend that should be further investigated in subsequent work.

In addition to these comments, note that the current survey did not incentivize voters to make good or bad decisions. Future directions could include testing whether the epistemic performance of liquid democracy and the roles of expertise and confidence change in the presence of rewards for good answers, taken either directly or through transitive delegations.

Next, our experiments, targeted to collect delegation behaviors, could not detect the significance of the use of liquid democracy itself as the group sizes were all relatively small. Further work is needed to test more challenging questions (where direct democracy may fall below 0.5) as well as the asymptotic epistemic performance that has been studied thus far in theoretical work [7, 16, 20].

Finally, note that the scope of this study is particularly narrow as it only considers binary questions with correct answers. In short, epistemic studies of voting relate primarily to the instrumental value of democracy. This informs efforts to deploy liquid democracy in prediction markets or to make corporate decisions with clear (but hard to achieve) goals. However, deploying liquid democracy in political settings, for instance, would require further research and tests on the intrinsic value of delegations and how they relate to paradigms of representation of conflicting moral values that may not reduce to factual evidence. Further work at the intersection of political philosophy and social choice would be most needed to understand these other aspects of liquid democracy.[6]

REFERENCES

[1] Kenneth J Arrow, Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D Nelson, et al. 2008. The promise of prediction markets. , 877–878 pages.

[2] Ruben Becker, Gianlorenzo D'Angelo, Esmaeil Delfaraz, and Hugo Gilbert. 2021. When Can Liquid Democracy Unveil the Truth? arXiv:2104.01828.

[3] Daan Bloembergen, Davide Grossi, and Martin Lackner. 2019. On rational delegations in liquid democracy. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI).* 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] Markus Brill and Nimrod Talmon. 2018. Pairwise Liquid Democracy.. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI).* 137–143.

[6] The two dimensions that are usually used to evaluate decision-making processes are the epistemic dimension (accuracy of the voting outcome.) and the procedural dimension (fairness of the procedure). Debates on the validity and comparison of the dimensions in institutional design are out of the scope of this paper, but interested readers should refer to Chiara Destri's essay on the matter [11].

[6] David V Budescu and Eva Chen. 2015. Identifying expertise to extract the wisdom of crowds. *Management Science* 61, 2 (2015), 267–280.
[7] Ioannis Caragiannis and Evi Micha. 2019. A Contribution to the Critique of Liquid Democracy. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*. 116–122.
[8] Zoé Christoff and Davide Grossi. 2017. Binary Voting with Delegable Proxy: An Analysis of Liquid Democracy. In *Proceedings of the 16th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*. 134–150.
[9] Nicolas De Condorcet. 1785. Essai sur l'application de l'analyse la probabilite des decisions rendues a la pluralite des voix. *Paris: L'Imprimerie Royale.* (1785).
[10] Dominik Dellermann, Nikolaus Lipusch, Philipp Ebel, and Jan Marco Leimeister. 2020. The potential of collective intelligence and crowdsourcing for opportunity creation. *International Journal of Entrepreneurial Venturing* 12, 2 (2020), 183–207.
[11] Chiara Destri. 2017. Right or wrong, it's democracy. Legitimacy, justification and the independent criterion. (2017).
[12] Francis Galton. 1907. Vox populi (the wisdom of crowds). *Nature* 75, 7 (1907), 450–451.
[13] Daniel G Goldstein, Randolph Preston McAfee, and Siddharth Suri. 2014. The wisdom of smaller, smarter crowds. In *Proceedings of the fifteenth ACM conference on Economics and computation*. 471–488.
[14] Paul Gölz, Anson Kahng, Simon Mackenzie, and Ariel D Procaccia. 2018. The fluid mechanics of liquid democracy. In *Proceedings of the 14th Conference on Web and Internet Economics (WINE)*. 188–202.
[15] James Green-Armytage. 2015. Direct voting and proxy voting. *Constitutional Political Economy* 26, 2 (2015), 190–220.
[16] Daniel Halpern, Joseph Y Halpern, Ali Jadbabaie, Elchanan Mossel, Ariel D Procaccia, and Manon Revel. 2021. In Defense of Liquid Democracy. arXiv:2107.11868.
[17] Stefan M Herzog and Ralph Hertwig. 2011. The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision making* 6, 1 (2011), 58–72.
[18] Shawndra Hill and Noah Ready-Campbell. 2011. Expert stock picker: the wisdom of (experts in) crowds. *International Journal of Electronic Commerce* 15, 3 (2011), 73–102.
[19] Karsten Hueffer, Miguel A Fonseca, Anthony Leiserowitz, and Karen M Taylor. 2013. The wisdom of crowds: Predicting a weather and climate-related event. (2013).
[20] Anson Kahng, Simon Mackenzie, and Ariel D Procaccia. 2021. Liquid democracy: An algorithmic perspective. *Journal of Artificial Intelligence Research* 70 (2021), 1223–1252.
[21] Abraham Kaplan, AL Skogstad, and Meyer A Girshick. 1950. The prediction of social and technological events. *Public Opinion Quarterly* 14, 1 (1950), 93–110.
[22] Craig A Kaplan. 2001. Collective intelligence: A new approach to stock price forecasting. In *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236)*, Vol. 5. IEEE, 2893–2898.
[23] Dimitri Landa and Ryan Pevnick. 2020. Representative democracy as defensible epistocracy. *American Political Science Review* 114, 1 (2020), 1–13.
[24] Hélène Landemore. 2012. *Democratic reason.* Princeton University Press.
[25] Hélène Landemore. 2020. *Open democracy.* Princeton University Press.
[26] Gustave Le Bon. 2009. *Psychology of Crowds (annotated).* Sparkling Books.
[27] Yiqun Liu, Fei Chen, Weize Kong, Huijia Yu, Min Zhang, Shaoping Ma, and Liyun Ru. 2012. Identifying web spam with the wisdom of the crowds. *ACM Transactions on the Web (TWEB)* 6, 1 (2012), 1–30.
[28] Charles Mackay. 1869. *Memoirs of extraordinary popular delusions and the madness of crowds.* George Routledge and sons.
[29] James Madison. 1787. The federalist no. 10. *November* 22, 1787 (1787), 1787–88.
[30] Tyler Moore and Richard Clayton. 2008. Evaluating the wisdom of crowds in assessing phishing websites. In *International Conference on Financial Cryptography and Data Security*. Springer, 16–30.
[31] Plato. 1911. *Plato's Phaedo.* Clarendon Press.
[32] Charles Polk, Robin Hanson, John Ledyard, and Takashi Ishikida. 2003. The policy analysis market: an electronic commerce application of a combinatorial information market. In *Proceedings of the 4th ACM Conference on Electronic Commerce*. 272–273.
[33] Drazen Prelec. 2004. A Bayesian truth serum for subjective data. *science* 306, 5695 (2004), 462–466.
[34] Manon Revel, Tao Lin, and Daniel Halpern. 2022. The Optimal Size of an Epistemic Congress. *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*.
[35] Joseph P Simmons, Leif D Nelson, Jeff Galak, and Shane Frederick. 2011. Intuitive biases in choice versus estimation: Implications for the wisdom of crowds. *Journal of Consumer Research* 38, 1 (2011), 1–15.
[36] Camelia Simoiu, Chiraag Sumanth, Alok Mysore, and Sharad Goel. 2019. Studying the "wisdom of crowds" at scale. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 171–179.
[37] James Surowiecki. 2005. *The wisdom of crowds.* Anchor.
[38] Philip E Tetlock and Dan Gardner. 2016. *Superforecasting: The art and science of prediction.* Random House.
[39] Chiara Valsangiacomo. 2021. Clarifying and Defining the Concept of Liquid Democracy. *Swiss Political Science Review* (2021).
[40] Chiara Valsangiacomo. 2021. Political representation in liquid democracy. *Frontiers in Political Science* (2021), 7.
[41] Thomas Walter and Andrea Back. 2010. Crowdsourcing as a Business Model: An exploration of emergent textbooks harnessing the wisdom of crowds. AIS.
[42] Yuzhe Zhang and Davide Grossi. 2021. Power in liquid democracy. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*. 5822–5830.

A SURVEY OUTLOOK

This section contains examples of the delegating and the voting pages.

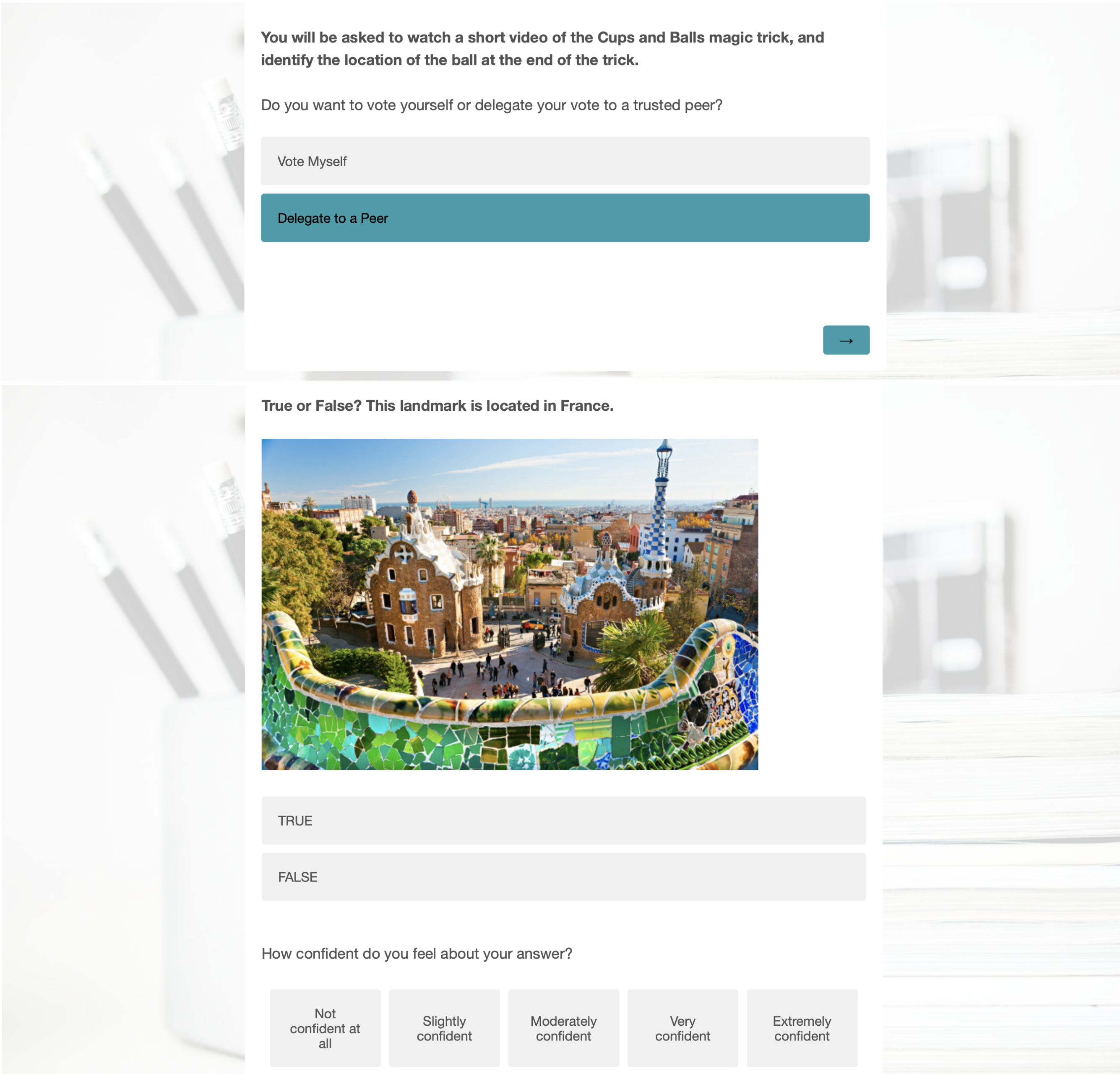

Figure 6: Excerpts from the Liquid Democracy Survey.

Example of survey task when participants were asked to delegate at the category-level (top) and to answer at a specific questions (bottom).