# Token-Weighted Crowdsourcing

Gerry Tsoukalas, Brett Hemenway Falk*

Sept 20, 2019

*Management Science*—forthcoming

**Abstract**

Blockchain-based platforms often rely on token-weighted voting ("$\tau$-weighting") to efficiently crowdsource information from their users for a wide range of applications, including content curation and on-chain governance. We examine the effectiveness of such decentralized platforms at harnessing the "wisdom" and "effort" of the crowd. We find that $\tau$-weighting generally discourages truthful voting, and erodes the platform's predictive power unless users are "strategic enough" to unravel the underlying aggregation mechanism. Platform accuracy decreases with the number of truthful users and the dispersion in their token holdings, and in many cases, platforms would be better off with a flat "$1/n$" mechanism. When, prior to voting, strategic users can exert effort to endogenously improve their signals, users with more tokens generally exert more effort—a feature often touted in marketing materials as a core advantage of $\tau$-weighting—however, this feature is not attributable to the mechanism itself, and more importantly, the ensuing equilibrium fails to achieve the first-best accuracy of a centralized platform. The optimality gap decreases as the distribution of tokens across users approaches a theoretical optimum, that we derive, but, tends to increase with the dispersion in users' token holdings.

*Keywords:* blockchain, crowdsourcing, cryptocurrency, information aggregation, on-chain governance, strategic voting, tokenomics, token-curated registries (TCR).

## 1 Introduction

Many blockchain-based platforms have implemented token-weighted voting ($\tau$-weighting) to incentivize efficient information crowdsourcing from their users in a decentralized way. These systems aggregate information through user votes, where the final action of the system is determined based on the weighted average of the users' votes, and each user's vote is weighted by his token holdings within the system. At their core, these systems work under the principle that users with more tokens have more "skin in the game" and are thus incentivized to provide "higher-quality" votes.

*Tsoukalas: Operations, Information and Decisions Department, University of Pennsylvania (gtsouk@wharton.upenn.edu); Falk: Department of Computer and Information Sciences, University of Pennsylvania (fbrett@cis.upenn.edu). The authors are thankful to the Ripple Research Fund at the Wharton School, for a grant that supported this work.

Though $\tau$-weighting has already been deployed on many live blockchain platforms, there is surprisingly little research attesting to its theoretical soundness. This paper examines some of its basic economics ("tokenomics") to gauge its effectiveness. In particular, we focus on the following questions: Does $\tau$-weighting encourage or discourage truthful voting? How "accurate" is the resulting crowdsourced information? Does having "skin in the game" in the form of token holdings adequately incentivize user effort to improve the platform's predictive power?

The surge of blockchain-based platforms over the past several years has brought with it many new challenges. One of the most critical issues faced by these platforms is the need to crowdsource information from their users. This issue is at the core of many decentralized systems, regardless of how different they may otherwise be. Fittingly, there is a wide range of applications ranging from governance issues such as how funds raised from an ICO (Initial Coin Offering) should be spent, to evaluating the quality of code-upgrade proposals, to soliciting user feedback on service experience, aggregating product quality ratings, combating fake news, and many others.

One of the first applications of $\tau$-weighted voting was in the "Slock.it" Decentralized Autonomous Organization (DAO) (Jentzsch 2016). The DAO was envisioned as a new type of user-managed investment fund, where users could deposit tokens into a joint fund, and the fund's investment decisions would be made by the weighted votes of the users. Although the Ethereum smart contract running the DAO was hacked, leading to its collapse, the DAO model inspired countless blockchain startups to adopt similar forms of on-chain governance, where token holders are allowed to direct the overall system either through informal (non-binding) votes like CarbonVote (Ashu and Lv 2018) or binding resolutions (Warren and Bandeali 2018, Eufemio et al. 2018, Goodman 2014). CarbonVote has tallied more than five million Ether towards votes (Ashu and Lv 2019), and Tezos had more than 28,000 votes in for its first, binding hard-fork proposal (Kim 2019).

Token-weighted voting has also been used for content curation where users can promote (upvote) content with their votes weighted proportionally to their token holdings—Steem (Steem 2018) and Sapien (Bhatia et al. 2018) are two notable examples. Steem identifies itself as a "blockchain-based rewards platform for publishers to monetize content," and has more than 40,000 daily active users (Arcange 2019). Sapien is a "social news platform that gives users control of their data, rewards content creators, and fights fake news." In a different application, a prototype implementing a variant of $\tau$-weighted voting has been developed to allow users to vote on the trustworthiness of TLS and SSL certificates (Hentschker 2018). A general model of "Token-Curated Registries" (TCRs) (Goldin 2017) has applied this principle to allow users to collectively curate a list of "high-quality" content. Civil (Iles 2018) has adopted the TCR model to combat fake news by identifying high-quality news sources, and AdChain (Goldin et al. 2017) uses a TCR to fight fraud in digital advertising.

Despite differences in their specific implementations, these systems all share a common underlying design and incentive mechanism: individual users are allowed to vote on the "quality" of content (e.g., a news source, social media post, an investment plan, the quality of a product or a service), and the system aggregates these votes, weighting them according to each user's token holdings. The system then takes action based on this weighted average. Importantly, the system's "reward" is proportional to the quality of the decision. This reward is realized in the form of increased token value, and is thus distributed among the users proportionally to their token holdings. The key design principle here is that users with large token holdings will be the most incentivized to increase the overall value of the platform.

These aforementioned features are at the core of our model of $\tau$-weighted crowdsourcing. We consider a product or a service of unknown quality subjected to a vote on a platform. The platform commits (e.g., through a smart contract) to aggregate user votes according to a standard $\tau$-weighted mechanism. Platform value is driven by accuracy, which depends on the nature of the incoming votes it receives from stakeholders. The more accurate the platform is, the more valuable its tokens, and the more value it generates for its stakeholders. The voters are simply the token holders of the platform and are heterogeneous along two dimensions: their token holdings and the precision of their private signals. We first examine a simple voting game with exogenous signal precisions, and consider two types of players: *strategic* and *truthful*. Truthful players simply vote their posterior beliefs after observing their own signal, while ignoring the presence of other voters on the platform. Strategic players, on the other hand, are fully rational—they report a vote that maximizes the expected value of their tokens, taking into account not only their own signal, but other voters' strategies as well. Next, recognizing that strategic players may want to try and improve their information prior to voting, we extend the base model to account for endogenous information acquisition, adding an additional stage of effort provisioning; that is, prior to choosing a voting strategy, each player has the option to exert (costly) effort to improve the precision of his own signal. We characterize the equilibria (in linear strategies) of the crowdsourcing games with exogenous and endogenous information acquisition.

In brief, our results suggest that when it comes to crowdsourcing information, $\tau$-weighted aggregation generally i) discourages truthful voting, and ii) reduces platform accuracy, unless players are sophisticated enough to endogenously unravel the underlying weighting mechanism. When we incorporate effort into our model to account for endogeneity in information acquisition, we show that $\tau$-weighting iii) provides some desirable "skin-in-the-game" type effort incentives for strategic players, but iv) nonetheless fails to achieve the first-best predictive power of a centralized platform.

Beyond these high-level insights, which provide answers to the questions raised earlier in the discussion, our analysis affords a series of more in-depth results on the nature and impact of strategic vs. truthful voting, the effectiveness of centralized vs. decentralized aggregation, and sensitivities to token dispersion (i.e., how tokens are distributed across users).

Intuitively, one might expect $\tau$-weighting to lead to suboptimal outcomes because it has the potential to exacerbate an inherent mismatch that may exist between the distribution of tokens and the distribution of information. For instance, a voter might have excellent information but low stake, in which case, his vote will be underweighted in the aggregate, and vice versa. We find that strategic voters are able to fully overcome this mismatch by purposely misreporting their beliefs to the platform, effectively unraveling the platform's "sub-optimal" weighting mechanism. However, this only holds when all voters are strategic. First-best accuracy is elusive when even a single truthful voter is present, and the optimality gap increases in the number of truthful voters. In such cases, we show the platform is generally better off with an un-weighted "$1/n$" aggregation mechanism. Taken together, these results suggest that if these decentralized platforms can effectively harness the wisdom of the crowd, it is *despite of, not because of,* their $\tau$-weighting mechanism.

To account for the fact that strategic voters may want to try and correct the aforementioned mismatch between the distribution of tokens and information, we then endogenize their information acquisition decisions in Section 4, where, prior to voting (stage 2), we give them the ability to improve the precision of their own private signal by exerting costly effort (stage 1). The extended two-stage model requires introducing two generic functions whose properties will be critical for the analysis: the information improvement function, which maps effort to signal accuracy, and the effort cost function, which maps effort to costs incurred.

We show that under certain technical/structural conditions on these functions, there is a unique effort-exerting equilibrium in which individual effort levels can be derived in closed form. Leveraging the expressions obtained, we show that effort increases with token holdings—a feature that platforms often tout in their marketing materials as a key advantage of $\tau$-weighted voting. We find the effect, however, not to be attributable to the token-weighting mechanism itself, which remains irrelevant when all users are strategic, even under endogenous information acquisition. Rather, it is simply due to the fact that players with more tokens stand to benefit more from an increase in platform accuracy, because accuracy directly drives token value. In other words, the same effect would be observed if the platform adopted a flat "$1/n$" aggregation mechanism, as long as agents with more tokens stand to benefit more from an increase in platform accuracy. We also show that ceteris paribus, an agent with higher precision will exert more effort than one with lower precision, implying some "free-riding" from less-informed voters.

Comparing the resulting equilibrium to that of a centralized platform that can coordinate user behavior, we find that decentralized equilibrium effort levels are strictly lower. As a result, the platform cannot achieve first-best effort provisioning in the decentralized setting, even in the best case scenario where all agents are strategic. We characterize the ensuing optimality gap and show that "on average" it grows with the dispersion in players' token holdings, implying platforms may generally prefer that tokens are not too disproportionately held by platform users. The effect is stronger when all players have homogeneous precisions, and weaker, but still positive, otherwise. Finally, we derive the platform's optimal allocation of tokens, as a function of the distribution of precisions across its users and show, through an appropriately chosen distance measure, that the platform would prefer dispersed distributions if these are "close enough" to the theoretical optimum, and a homogeneous distribution otherwise.

To summarize, our results bring to light some of the more subtle pros and cons of $\tau$-weighted aggregation systems, and, in contrast to what their wide-spread use implies, raise some questions about their effectiveness at harnessing information and user effort. Despite the ever-increasing popularity of $\tau$-weighted aggregation systems in practice, the academic literature has remained relatively silent on their theoretical soundness. This paper, which can be seen as a step to help bridge this gap, strives to put the emerging topic of $\tau$-weighted aggregation on a firmer foundation, and to provide some guidance on the design of blockchain-based voting systems.

## Related Literature

Though we are not aware of any other theoretical papers focusing specifically on the effectiveness of $\tau$-weighted crowdsourcing, our work is related to several literature streams.

*Blockchain systems & Token-Curated Registries (TCRs).* Blockchain-based systems provide a natural platform for $\tau$-weighted voting, since each player's token holdings are usually publicly known (on the blockchain) and ballots can be cast, aggregated and verified in a completely automated manner. In fact, numerous $\tau$-weighted voting mechanisms have already been deployed on different blockchain platforms as mentioned in the introduction. There is a growing practitioner literature on the topic of token-curated registries (Goldin 2017), however, the studies focus on practical implementation (e.g., coding details) and do not seek to formally model strategic agent behavior, like we do here. On the academic side, some studies have recently started to appear, for instance, Asgaonkar and Krishnamachari (2018) dive into some of the technical details of TCRs in a deterministic setting with known information. Many variants of TCRs have been proposed (Simon 2017, Lockyer 2018), and we expect these platforms to continue to grow going forward. Abstracting away from the technicalities of these systems, our study sheds light on the effectiveness of the token weighting mechanism that often sits at their core.

More broadly, the paper contributes to a rapidly growing literature discussing economic incentives in blockchain systems (Biais et al. 2017, Cong et al. 2018, Saleh 2018, Hinzen et al. 2019, Rosu and Saleh 2019), and studying the implications of the technology for a variety of areas such as auditing (Cao et al. 2018), corporate governance (Yermack 2017), crowdfunding (Chod and Lyandres 2018, Gan et al. 2019), finance (Biais et al. 2018), innovation (Catalini and Gans 2017), operations management and supply chains (Babich and Hilary 2018, Chod et al. 2018), etc.

*Crowdsourcing & Information Sharing in Networks.* Our work is closely related to the literature on crowdsourcing, which studies the ability of firms to source information (Araman and Caldentey 2016, Papanastasiou et al. 2017), funds (Alaei et al. 2016, Strausz 2017, Babich et al. 2019, Belavina et al. 2019), or innovation (Terwiesch and Xu 2008, Bimpikis et al. 2015, Stouras et al. 2017) from users. We are not aware of any work in this literature that like us, examines the feasibility and effectiveness of crowdsourcing information and effort using a $\tau$-weighted mechanism.

More broadly, our work is related to the literature on information sharing in networks which studies information exchange and aggregation, either through direct communication (Acemoglu et al. 2014) or through observational learning (Acemoglu et al. 2011). In a related but different setting, Saghafian et al. (2018) examine information aggregation through (imperfect) sensors that can solicit information from other sensors. Unlike us, these papers do not focus on the strategic incentives that agents have to (mis)report their information and exert effort when subjected to a $\tau$-weighted voting mechanism.

*Weighted Voting & Shareholder Voting.* Our work is broadly related to the literature on weighted voting mechanisms which have been studied in a variety of different settings. For example, Banzhaf III (1964) and a series of papers that build on this work (see e.g., Snyder Jr et al. (2005) and references therein) examine electoral weighted voting, focusing on the distinction between voting rights and voting power. Taylor and Zwicker (1993), Nordmann and Pham (1999), and Elkind et al. (2008) focus on coalition structures and the division of power in weighted voting games. Gifford (1979) and Tong and Kain (1991) study weighted voting in distributed computing systems focusing on characterizing computational complexity.

Weighted voting is also relevant for the literature on shareholder voting, given that one share often entitles one vote, and shareholders hold different share amounts. Papers in this area tend to focus on issues specific to the context, e.g., the mismatch of incentives between managers and shareholders (Shleifer and Vishny 1986), issues of vote trading (Christoffersen et al. 2007), and the differences between binding vs. non-binding votes (Levit and Malenko 2011).

Our work departs from these settings in a number of ways. First, the input and/or output space studied in these papers is generally constrained to be discrete, and in most cases, binary (yes/no). With these constraints, the complexity comes from the combinatorial nature of the problem. In contrast, the choice of continuous input and output spaces in our model deliberatly draws from the established economics and finance literatures on information aggregation/acquisition (Myatt and Wallace 2012, Colombo et al. 2014), and price informativeness (Vives 1988, 2011, Ostrovsky 2012, Even et al. 2019), and is meant to capture situations that are best approximated by continuous states. For instance, our model is more suitable for situations in which agents are asked to estimate quality of content or products, as in the TCR examples mentioned previously, and/or investment amounts, as in the DAO example, and less suitable to inform on yes/no issues of governance, such as electing board members, stock splits, M&A decisions, etc. Second, papers in these areas generally do not focus on user effort incentives, which is one of the main contributions of our work, and one of the main areas of focus of $\tau$-weighted platforms in their marketing materials. Third, these papers do not specifically focus on the tokenomics of $\tau$-weighted crowdsourcing systems. As such, they do not seek to address the specific questions we examine, including truthful vs. strategic voting, centralized vs. decentralized aggregation, $\tau$-weighting vs. $1/n$-weighting, token dispersion, and the implications these have for token value and platform effectiveness.

## 2 Model

Consider a product (or a service) of unknown quality, $q$, subject to a vote on a platform with $n$ token holders ("players"), indexed by $i \in \{1, \ldots, n\}$. Players cannot directly observe $q$, but have a common Gaussian prior belief

$$q \sim \mathcal{N}(\mu, \sigma_q^2),$$

where $\mathbb{E}[q] = \mu$ is the mean quality and $\mathrm{Var}[q] = \sigma_q^2$ captures quality dispersion, which can be a proxy for the novelty of the product.[1]

Players are tasked to vote on quality and are heterogeneous along two dimensions: Their relative

[1]For instance, when $\sigma_q \to 0$, there is no uncertainty about product quality. Conversely, when $\sigma_q \to \infty$, any value of $q$ on the real line is equally likely, e.g., this could represent an innovative product that has just hit the market. Formally, $\sigma_q \to \infty$ represents the case of a uniform improper prior.

token holdings $\tau_i \in (0,1)$, normalized so that $\sum_{i=1}^{n} \tau_i = 1$, and the signal they obtain about the product quality. More specifically, player $i$ receives a private noisy signal

$$s_i = q + \epsilon_i, \tag{1}$$

where $\epsilon_i$ is a normally distributed noise term with $\mathbb{E}[\epsilon_i] = 0$ and $\mathrm{Var}[\epsilon_i] = \sigma_i^2$, i.e., $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2), \forall i \in \{1, \ldots, n\}$. It follows that signals are independent and normally distributed with $s_i \sim \mathcal{N}(\mu, \sigma_q^2 + \sigma_i^2)$. In line with extant literature, $\sigma_i, i \in \{1, \ldots, n\}$, are publicly visible, but players cannot observe others' private signals. Similarly to how these platforms operate in practice, token holdings $\tau_i, i \in \{1, \ldots, n\}$, are also publicly visible.[2]

After observing their private signal, players simultaneously submit their "votes" $v_i = v_i(s_i)$; in line with extant literature (e.g., Vives 1988, Myatt and Wallace 2012), we restrict our attention to the class of linear strategies, represented by letter $\mathbb{L}$, i.e.,

$$v_i = \alpha_i s_i + (1 - \alpha_i)\mu, \tag{2}$$

where $\alpha_i \in \mathbb{R}$ is the "weight" player $i$ chooses to place on his signal.[3] We note that we purposely do *not* restrict $\alpha_i$ to the interval $[0, 1]$, so assuming $v_i$ to be a linear combination of $s_i$ and $\mu$ does not restrict the range of $v_i$. The linear model also facilitates comparisons with Bayesian posteriors which are also linear in the signal. In particular, conditional on $s_i$, $q$ is normally distributed with mean $\mathbb{E}[q \mid s_i] = \beta_i s_i + (1 - \beta_i)\mu$, where $\beta_i = \sigma_q^2 \left(\sigma_q^2 + \sigma_i^2\right)^{-1}$ is the "weight" player $i$ places on his signal, and $\mathrm{Var}[q \mid s_i] = \left(\sigma_q^{-2} + \sigma_i^{-2}\right)^{-1}$.

Later, in Section 4, we will endow users with the ability to exert effort to improve their signals, but we defer the modeling extension required to that section.

The platform aggregates incoming votes weighted by token holdings, which forms the platform's quality estimate $\hat{q}$, that is,

$$\hat{q}(v_1, \ldots, v_n) = \sum_{i=1}^{n} \tau_i v_i. \tag{3}$$

Consistent with how these platforms operate in practice, the platform commits to aggregation mechanism (3), e.g., through an auditable "smart contract." Throughout the paper we will compare this mechanism, which we refer to as *$\tau$-weighted aggregation*, or simply $\tau$-weighting, to a benchmark

[2]Relaxing these basic assumptions generally has non-trivial consequences, and could thus be an interesting direction for future research.

[3]It is known that when all other players are playing linear strategies, the focal agent's best response is linear (see e.g., Myatt and Wallace (2012) pp. 347). This assumption is often adopted in the literature to ensure tractability.

of an equally weighted (flat) mechanism, which we refer to as $1/n$-*weighted aggregation*, or simply $1/n$-weighting, that assumes all players are attributed a weight $1/n$, irrespective of their token holdings; formally $\hat{q}_{1/n}(v_1, \ldots, v_n) = \frac{1}{n} \sum_{i=1}^{n} v_i$. For the sake of completeness, we will also consider the case of a platform optimizing the aggregation mechanism in the Appendix, Section A.3.

Players are rewarded through the value of their tokens, which in turn, depends on the accuracy of the platform.[4] In particular, the closer the aggregate quality, $\hat{q}$, is to $q$, the more valuable the platform becomes, and hence, the more valuable the tokens become.

To model the platform's value, we will work with a general class of differentiable and concave payoff functions taken over the error $(q - \hat{q})$, and will assume these to be "well-behaved" if they satisfy the following properties.

**Assumption 1** (Well-Behaved Payoffs)**.** *A payoff function* $\pi : \mathbb{R} \to \mathbb{R}$ *is said to be well-behaved if: (1)* $\pi$ *is symmetric about the origin,* i.e., $\pi(x) = \pi(-x)$*, and (2)* $\pi$ *is decreasing away from the origin,* i.e., $\pi(x) \leq \pi(y)$ *whenever* $|x| \geq |y|$*.*

The "well-behaved" assumption means that over-estimating and under-estimating the quality are equally bad, and that the more accurate the estimate, that is, the lower the error $(q - \hat{q})$, the higher the payoff. As we shall see, these basic assumptions suffice to obtain meaningful results in Section 3, without having to restrict the payoff to a specific functional form. Many standard functional forms would satisfy these conditions, for instance, $\pi(x) \propto -x^2$, which would correspond to a quadratic utility, or $\pi(x) \propto e^{-x^2}$, etc.

Defining the platform's payoff as $\pi(q - \hat{q})$, we can also define each player's individual share of this payoff by $\pi_i = \tau_i \cdot \pi$, i.e., consistent with how these platforms operate in practice, each player obtains a share of the payoff proportional to his relative share of the tokens. This feature drives several of our results, as we shall see in Section 4.

Our core model assumes all players are strategic (fully rational), represented by letter $\mathbb{S}$, and we seek to characterize the linear pure-strategy Bayesian Nash equilibria of the game. More specifically, each player $i$ maximizes his expected payoff over his linear voting strategy $v_i \in \mathbb{L}$, conditional on the private signal he receives, $s_i$, given the platform aggregation mechanism (3), and the linear

[4] While token value may be driven by additional factors, such as the secondary market liquidity of the tokens, the quality of the underlying blockchain sustaining the tokens, the presence of speculators in the market, etc., our primary interest in this paper is to assess whether these platforms can effectively crowdsource information solely based on accuracy incentives. We therefore consider accuracy as the primary driver of token value, and by implication, platform value.

strategies played by others, $\boldsymbol{v}_{-i} \in \mathbb{L}$,

$$\max_{v_i \in \mathbb{L}} \mathbb{E}\left[\pi_i\left(q - \hat{q}(v_1, \ldots, v_n)\right) \mid s_i\right]. \tag{4}$$

In Section 3, the platform composed of strategic players will be compared to a benchmark platform consisting of non-strategic players following simple "truthful" voting strategies. Truthful voters, represented by letter "$\mathbb{T}$," vote their true, albeit individual, Bayesian posteriors. That is, they update their beliefs based on their own private signal, setting weight $\alpha_i = \beta_i = \sigma_q^2\left(\sigma_q^2 + \sigma_i^2\right)^{-1}$ and vote $v_i = \beta_i s_i + (1-\beta_i)\mu$, without considering other players. Practically speaking, platforms may be composed of players exhibiting different degrees of strategic behavior (another interpretation is that players may have different behavioral biases), hence, it is meaningful to examine the implications that player heterogeneity along this dimension can have on platform accuracy. The above two player types we consider can be thought of as coarse representation of the spectrum.[5] In Section 3.3, we extend the model to the more realistic case where both player types are simultaneously present on the platform.

# 3 Crowdsourcing Information

In this section, we examine equilibrium voting strategies and their implications for the platform's predictive accuracy. We begin with the first-best variance obtainable in the centralized setting.

## 3.1 First-Best Platform Variance

Before presenting the result, we introduce one intermediate technical lemma that will be useful throughout the analysis.

**Lemma 1** (Payoff-Variance Equivalence)**.** *Suppose $\pi : \mathbb{R} \to \mathbb{R}$ satisfies Assumption 1 and let $N(0, \sigma^2)$ denote a normally distributed random variable with mean 0 and variance $\sigma^2$, then*

*(i)* $\mathbb{E}\left[\pi(N(0,\sigma^2))\right]$ *is monotonically decreasing in* $\sigma^2$*;*

*(ii)* $\mathbb{E}\left[\pi(N(0,\sigma^2))\right] \geq \mathbb{E}\left[\pi(N(x_0,\sigma^2))\right]$ *for all* $x_0$*;*

*(iii) if* $V(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}^+$*,* $S \subset \mathbb{R}^n$ *is a set of constraints, and* $\boldsymbol{x}^* \stackrel{\text{def}}{=} \underset{\boldsymbol{x} \in S}{\operatorname{argmin}} V(\boldsymbol{x})$*, then* $\mathbb{E}\left[\pi\left(N(0, V(\boldsymbol{x}^*))\right)\right] = \max_{\boldsymbol{x} \in S} \mathbb{E}\left[\pi\left(N(0, V(\boldsymbol{x}))\right]\right.$.

[5] Note, an additional, even less sophisticated truthful player type that simply votes his "raw" signal, $v_i = s_i$, instead of his Bayesian posterior, is subsumed in the current model by taking the limiting case of a diffuse prior, $\sigma_q \to \infty$, which implies $\beta_i \to 1$, and hence $v_i \to s_i$.

Proofs are provided in the Appendix. One useful implication of Lemma 1 is that since the payoff function is well-behaved and the aggregate error $(q - \hat{q})$ is normally distributed with mean 0, we can interchangeably talk about the platform/players maximizing expected payoff, or minimizing the variance of the aggregate error.

Consider an omniscient central planner that can directly observe all signals and set $\hat{q}$ to maximize $\mathbb{E}\left[\pi(q - \hat{q}) \mid s_1, \ldots, s_n\right]$.[6] In light of Lemma 1, this can equivalently be written as a minimization problem over the variance, $V^{fb} \stackrel{\text{def}}{=} \min \operatorname{Var}\left[q - \hat{q} \mid s_1, \ldots, s_n\right] = \operatorname{Var}\left[q \mid s_1, \ldots, s_n\right]$. The latter is simply the Bayesian posterior variance in the standard multivariate Gaussian setting,[7] thus

$$V^{fb} = \left(\sigma_q^{-2} + \sum_i^n \sigma_i^{-2}\right)^{-1}. \tag{5}$$

As expected, $V^{fb}$ increases in $\sigma_q$ and $\sigma_i$, i.e, the more imprecise the prior and signals are, the worse is the platform's predictive power over $q$. Furthermore, the platform becomes perfectly accurate as the number of players $n \to \infty$. In other words, the centralized platform can fully harness the wisdom of the crowd at the limit. This, however, hinges on "reasonable" $\sigma_i, \forall i$, e.g., independence of individual signals and of their variance from $n$. For an alternative setting in which aggregate information does not necessarily grow with the number of users, see Bergemann and Välimäki (1997).

Finally, as is standard in the literature, it will be convenient to formally define the platform's "first-best precision" as the inverse of its first-best variance, $\rho^{fb} = \left(V^{fb}\right)^{-1}$. Combining this definition with the fact that the platform's estimate is unbiased (given all signals are unbiased and the aggregation mechanism is linear), we can say that the centralized setting yields the first-best (highest possible) accuracy the platform can achieve, and use the terms "precision" and "accuracy" interchangeably.

## 3.2 Equilibrium with Strategic Players

Next, we examine player voting strategies and the resulting platform accuracies they can achieve, and then compare these to the first-best benchmark.

To build intuition, consider first the simple case of a platform consisting of a single strategic player, holding all tokens (*i.e.,* $\tau_1 = 1$). In this case, $\hat{q} = \tau_1 \cdot v_1 = v_1$ and the player's optimal

[6] Alternatively, we can consider a central planer that observes all signals and sets player votes $v_i$ instead of directly setting $\hat{q}$, and the same results will hold.

[7] To see this, first note that the optimal $\hat{q}^*$ solves $\mathbb{E}\left[q - \hat{q}^* \mid s_1, \ldots, s_n\right] = 0$, which gives $\hat{q}^* = \mathbb{E}\left[q \mid s_1, \ldots, s_n\right]$. As a result, $\operatorname{Var}\left[q - \hat{q}^* \mid s_1, \ldots, s_n\right] = \operatorname{Var}\left[q - \mathbb{E}\left[q \mid s_1, \ldots, s_n\right] \mid s_1, \ldots, s_n\right] = \operatorname{Var}\left[q \mid s_1, \ldots, s_n\right]$. Since $q$ and $\{s_i\}_{i=1}^n$ are normally distributed one can use the standard result concerning multivariate normals $\operatorname{Var}\left[q \mid s_1, \ldots, s_n\right] = \sigma_q^2 - \mathbf{\Sigma}_{q\vec{s}} \mathbf{\Sigma}_{\vec{s}\vec{s}}^{-1} \mathbf{\Sigma}_{\vec{s}q}$, where $\mathbf{\Sigma}_{q\vec{s}}$, $\mathbf{\Sigma}_{\vec{s}\vec{s}}$ and $\mathbf{\Sigma}_{\vec{s}q}$ are the covariance matrices. A straightforward calculation gives (5).

strategy, conditional on observing signal $s_1$, is simply to truthfully report his Bayesian posterior, $v_1^* = \mathbb{E}[q \mid s_1] = \beta_1 s_1 + (1-\beta_1)\mu$, where $\beta_1 = \sigma_q^2(\sigma_q^2 + \sigma_1^2)^{-1}$ (see Appendix A.2 for details). We emphasize two points: first, the importance of Assumption 1. In Appendix A.1 we give some natural payoff functions (that are not well-behaved, *i.e.,* violate Assumption 1), where the player's optimal strategy is *not* to vote his Bayesian posterior. One particularly illustrative case that we discuss is that of a "price-is-right" type of payoff function.[8] Second, though truthfully voting one's posterior is intuitive, it does not carry over to the general $n$-player game, which we describe next.

**Proposition 1** (Voting Equilibrium)**.** *If the payoff function satisfies Assumption 1, and all players are strategic, then in equilibrium i) player $i$ votes $v_i^* = \alpha_i^* s_i + (1-\alpha_i^*)\mu$, where $\alpha_i^*$ is given in* (6) *and ii) the platform achieves the first-best variance given by* (5)*.*

$$\alpha_i^* = \frac{\sigma_q^2}{\tau_i \sigma_i^2 \left(1 + \sum_{j=1}^n \frac{\sigma_q^2}{\sigma_j^2}\right)}, \quad i \in \{1, \dots, n\}. \tag{6}$$

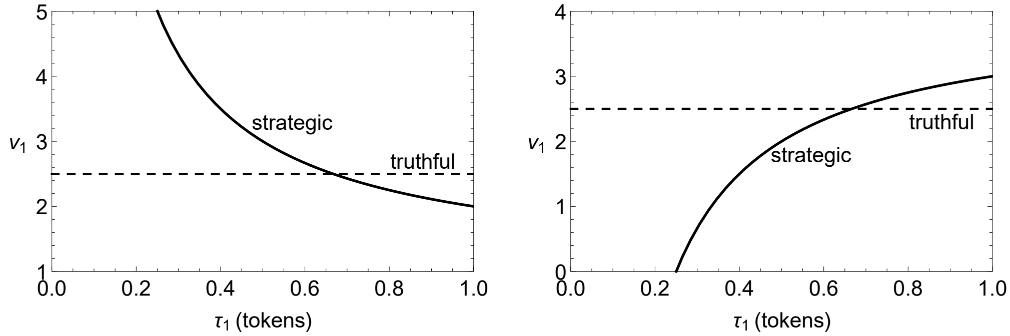
We outline three main takeways from Proposition 1. First, strategic players are able to recover first-best optimality in token-weighted platforms, but to do so, they must be willing to adjust votes based on the presence of their peers on the platform and their *own* token holdings. The former is observable from the dependence of the optimal weight on $\sigma_i, \forall i$, and the latter from its inverse relationship to $\tau_i$ (and its independence from $\tau_j, j \neq i$).

Second, the aggregation mechanism is irrelevant when the platform is composed of strategic players, that is, first best can also be restored under $1/n$-weighted aggregation (this is best seen by going through the proof of Proposition 1, replacing $\tau_i$ by $1/n, \forall i$).

Third, the Bayesian weight $\beta_i \neq \alpha_i^*$. This implies that truthful voting is generally suboptimal under $\tau$-weighted aggregation. Put differently, strategic players' votes will not reflect their true individual posterior beliefs. The votes cast can be either higher (vote inflation) or lower (vote shading) compared to truthful votes. To illustrate, Figure 1 compares player 1's equilibrium vote upon receipt of a positive ($s_1 = 4 > \mu = 1$ in Figure 1(a)), or negative ($s_1 = 1 < \mu = 4$ in Figure 1(b)) signal, in a $n = 2$-player platform.

This type of strategic behavior may not be harmless in practice. Consider for instance the implications for rating and review platforms. If individual ratings are visible (as opposed to just aggregate scores), customers who rely on such platforms for information may be exposed to a dis-

[8] In the popular U.S. game show "The Price is Right," which has been running since 1972, contestants (voters in our framework) compete to guess the price of an item, but face the prospect of elimination if their guess ends up being above the true price. We show that simple Bayesian posterior voting is generally suboptimal when it comes to dealing with this type of asymmetry in player payoffs.




(a) $v_1$ after positive signal

(b) $v_1$ after negative signal

Figure 1: Player 1's vote $v_1$ upon receiving a signal $s_1$, vs. his relative token holdings $\tau_1$, in a 2-player platform. Parameters $\sigma_1 = \sigma_2 = \sigma_q = 1/2$.

proportionate amount of extreme ratings, either very good, or very bad scores. For example, in Figure 1, strategic players with few tokens ($\sim$0.2) who receive a positive (negative) signal, inflate (deflate) their ratings well beyond their true beliefs. Without diving into a discussion on behavioral foundations, this type of outcome could possibly undermine the credibility of the platform as customers may fail to realize that this "gaming" of ratings is actually in their best interest, in terms of achieving maximal aggregate accuracy.

How realistic it is to assume all voters are strategic in practice, and are thus capable of restoring first best, is debatable. If some voters are in fact less sophisticated, for instance, if they have an inherent preference to report their true beliefs, two sets of questions arise. First: Q1) can strategic players restore first best in the presence of truthful voters, and if not, Q2) how does the resulting optimality gap vary with the relative number of truthful voters? Second: Q3) In the presence of truthful voters, should the platform prefer a $\tau$-weighted or equally weighted mechanism? Q4) Relatedly, how is this preference affected by the distribution of tokens across voters?

We address the first set of questions in Section 3.3, and the second set in Section 3.4.

### 3.3 Equilibrium with Mixed Player Types

We extend the model from Section 2 to consider a platform containing a mixture of $|\mathbb{T}|$ truthful and $|\mathbb{S}|$ strategic players, such that $|\mathbb{T}| + |\mathbb{S}| = n$ (notation $|\cdot|$ represents set cardinality). As before, strategic players cannot observe others players' signals or votes, but they can observe truthful voters' identities.[9] Truthful players, by definition, only consider their own signals. Before stating the result,

[9]This assumption, made for tractability, eliminates the need to define additional belief sets but begs the question of how player types could be revealed in practice; a potentially interesting direction for future work. See Chandrasekhar et al. (2015) for a related experimental study.

let weights $\alpha_i^{\mathbb{S}} \overset{\text{def}}{=} \alpha_i^*, i \in \mathbb{S}$, from (6), where the summation 1 to $n$ in the denominator is now over the set of strategic players $\mathbb{S}$.

**Proposition 2** (Voting Equilibrium with Mixed Types)**.** *In the presence of truthful voters on the platform, if the payoff function satisfies Assumption 1, then in equilibrium i) strategic player i votes* $v_i^{\mathbb{T},\mathbb{S}} = \alpha_i^{\mathbb{T},\mathbb{S}} s_i + (1 - \alpha_i^{\mathbb{T},\mathbb{S}})\mu$, *where* $\alpha_i^{\mathbb{T},\mathbb{S}} = \alpha_i^{\mathbb{S}} \left(1 - \sum_{j \in \mathbb{T}} \tau_j \beta_j\right), i \in \mathbb{S}$, *and ii) the corresponding platform variance* $V^{\mathbb{T},\mathbb{S}} = \left(1 - \sum_{j \in \mathbb{T}} \tau_j \beta_j\right)^2 V^{fb} + \sum_{j \in \mathbb{T}} (\tau_j \beta_j \sigma_j)^2 > V^{fb}$.

Proposition 2 shows that in this more realistic setting, there is still a unique equilibrium in which strategic players adjust their optimal signal weights from (6) to try and correct for the presence of truthful voters. However, in contrast to the result in Proposition 1, the platform no longer achieves first-best predictive accuracy. Interestingly, even the presence of a single truthful voter ($|\mathbb{T}| = 1$) prevents strategic players from restoring first best under $\tau$-weighted aggregation, and the more truthful players are present, the worse the platform accuracy becomes. Proposition 3, below, formalizes the latter statement.

**Proposition 3** (Platform Accuracy with Mixed Types)**.** *Consider a platform with a partition* $|\mathbb{S}|, |\mathbb{T}|$ *of a set of* $n$ *total players. For fixed* $n$, *if any number* $k \in \{1, \ldots, |\mathbb{T}|\}$ *of truthful players become strategic, then the platform's accuracy increases.*

As a reminder, the results hold assuming the platform has pre-committed to a token-weighted (or equally weighted) aggregation mechanism, e.g., via a smart contract. For the sake of completeness, we relax this assumption in Appendix A.3, where we assume the platform can optimally set the weights it attributes to each player's incoming vote. With this type of precise control, we show that the platform would indeed be able to recover first-best variance as long as it knows the identity of truthful voters. Note, however, that centralizing the aggregation mechanism in this way would go against the core philosophy these platforms were built on—that of decentralizing decisions — so may be of limited applicability.

### 3.4 Properties of the Optimality Gap

Having shown that the presence of truthful players generates an indelible optimality gap in Section 3.3 (answering questions Q1 and Q2), we now examine platform preference between $\tau$- and $1/n$-weighting (question Q3), and the effects of token dispersion (question Q4). To ease exposition, for the rest of this subsection we will assume the "worst case" scenario of a platform consisting entirely of truthful players, $|\mathbb{T}| = n$. Our qualitative insights remain valid for $1 < |\mathbb{T}| < n$ (for

$|\mathbb{T}| = 0$, there is no optimality gap as per Proposition 1). Before proceeding with the analysis, we introduce some necessary notation and intermediate results.

*Platform Variance.* Let $V_j^i, i \in \{\mathbb{T}, \mathbb{S}\}, j \in \{\tau, 1/n\}$, represent the variance achieved by the platform under truthful ($i = \mathbb{T}$) or strategic ($i = \mathbb{S}$) voting, and under $\tau$-weighting ($j = \tau$) or $1/n$-weighting ($j = 1/n$). Note that, from Proposition 1, $V_\tau^{\mathbb{S}} = V_{1/n}^{\mathbb{S}} = V^{\mathbb{S}} = V^{fb}$, that is, strategic players achieve first best under both mechanisms. The corresponding values are in Table 1.

Table 1: Player voting strategies and platform variance.

| | **Vote** | **Platform Variance** |
|---|---|---|
| Truthful $\tau$-weight. | $v_i = \beta_i s_i + (1 - \beta_i)\mu$ | $V_\tau^{\mathbb{T}} = \sum_{i=1}^n (\tau_i \beta_i \sigma_i)^2 + \sigma_q^2 \left(1 - \sum_{i=1}^n \tau_i \beta_i\right)^2$ |
| Truthful $1/n$-weight. | $v_i = \beta_i s_i + (1 - \beta_i)\mu$ | $V_{1/n}^{\mathbb{T}} = \sum_{i=1}^n \left(\frac{1}{n}\beta_i \sigma_i\right)^2 + \sigma_q^2 \left(1 - \sum_{i=1}^n \frac{1}{n}\beta_i\right)^2$ |
| Strategic | $v_i = \alpha_i^* s_i + (1 - \alpha_i^*)\mu$ | $V^{\mathbb{S}} = V^{fb} = \left(\sigma_q^{-2} + \sum_{i=1}^n \sigma_i^{-2}\right)^{-1}$ |

Finally, define the optimality gap between truthful and strategic (first best) voting as follows.

$$\mathcal{G}_j = V_j^{\mathbb{T}} - V^{fb} = V_j^{\mathbb{T}} - V^{\mathbb{S}}, \quad j \in \{\tau, 1/n\}.$$

*Token Dispersion.* Define $d(\boldsymbol{x}, \boldsymbol{y})$ as the Euclidean distance ($L^2$ norm) between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) = ||\boldsymbol{x} - \boldsymbol{y}||_2 = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$. Let the $n$-dimensional vectors $\boldsymbol{\tau} \stackrel{\text{def}}{=} \{\tau_1, \ldots, \tau_n\}$, $\boldsymbol{n}^{-1} \stackrel{\text{def}}{=} \left\{\frac{1}{n}, \ldots, \frac{1}{n}\right\}$ and $\mathbf{1} \stackrel{\text{def}}{=} \{1, \ldots, 1\}$. Lastly, define the dispersion of vector $\boldsymbol{\tau}$ as the Euclidean ($L^2$) norm from its average value, i.e.,

$$\text{disp}(\boldsymbol{\tau}) = d(\boldsymbol{\tau}, \bar{\tau}\mathbf{1}) = ||\boldsymbol{\tau} - \bar{\tau}\mathbf{1}||_2, \text{ where } \bar{\tau} = \frac{1}{n}\sum_{i=1}^n \tau_i = \frac{1}{n}.$$

With notation and definitions in place, we proceed with the analysis focusing first on the simple case in which all players have homogeneous signal precisions, before turning to the more general case with precision heterogneneity.

### Homogeneous precisions ($\sigma_i = \sigma, \ \forall i$)

Within this case of homogeneous precisions, we also distinguish between homogeneous and dispersed token holdings. The fully homogeneous case is relatively straightforward. Given there is no token dispersion in this case, there is no difference between $\tau$- and $1/n$-weighting. Simplifying the expressions in Table 1 we obtain

$$\mathcal{G}_\tau = \mathcal{G}_{1/n} = \frac{(n-1)^2\sigma^4\sigma_q^4}{n(\sigma^2+\sigma_q^2)^2(\sigma^2+n\sigma_q^2)} \geq 0,$$

which is decreasing in $n$. As a side remark, as $n \to \infty$, $\mathcal{G}_\tau \to \frac{\sigma^4\sigma_q^2}{(\sigma^2+\sigma_q^2)^2} \geq 0$, that is, though the optimality gap is decreasing in the number of players, it can persist at the limit as $n$ tends to infinity, meaning, neither aggregation scheme can fully harness the wisdom of a truthful voting crowd, no matter how large.[10]

Next, we consider heterogeneity in token holdings while keeping all players' signals equally precise.

**Proposition 4** (Dispersed Tokens)**.** *When players have heterogeneous token holdings and homogeneous signal variances,*

*(i)* $\mathcal{G}_\tau - \mathcal{G}_{1/n} \geq 0$*, i.e., 1/n-weighting dominates token-weighting;*

*(ii)* $\mathcal{G}_\tau - \mathcal{G}_{1/n}$ *increases in* $\mathrm{disp}(\boldsymbol{\tau})$*, i.e.,* $1/n$*-weighting dominance increases with token dispersion.*

In line with intuition, when all players have the same information precision, there is nothing to be gained by weighing one player's vote more or less than any other's. Thus, when the platform is composed of truthful, equally informed players, it is better off adopting an equally weighted mechanism and/or reducing token dispersion, if possible.

**Heterogeneous precisions**

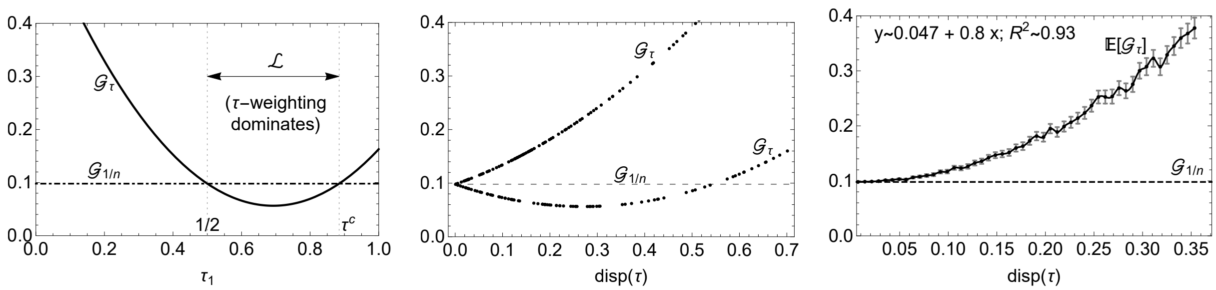
Here, we consider heterogeneity in both token holdings and signal variances, and examine the impact of token dispersion on the optimality gap. In this most general case, the comparison between $\tau$-weighting and $1/n-$weighting is non-trivial and either aggregation mechanism could a priori outperform. Intuitively, which mechanism "wins" depends on the mismatch that may exist between the distribution of tokens and the distribution of information. For instance, a voter might have excellent information but low stake, in which case, his vote will be underweighted in the aggregate, and vice versa. As we shall see, this intuition is only partially true.

The set of $\boldsymbol{\tau}$'s for which $\tau$-weighting dominates is given by the sublevel set of $V_\tau^{\mathbb{T}}$, $\mathcal{L} \stackrel{\text{def}}{=} \{\boldsymbol{\tau} \,|\, V_\tau^{\mathbb{T}}(\boldsymbol{\tau}) \leq V_{1/n}^{\mathbb{T}} \cap \mathbf{1}'\boldsymbol{\tau} = 1\}$. This set generally needs to be computed numerically, with the exception of low values of $n$. We focus here on the simplest case $n = 2$ which suffices to extract

[10] To see why, consider that as $n \to \infty$, the optimal weight to place on the prior $1 - \alpha^* = 1 - n/\left(n + \frac{\sigma^2}{\sigma_q^2}\right) \to 0$, whereas truthful voters maintain a fixed weight of $1 - \beta > 0$, which by definition, is independent of $n$.

the main qualitative insights. The general case $n > 2$ is analyzed in Appendix §A.4, and confirms result robustness.

Consider a platform with two players, and assume player one is "better informed" than player two, i.e., $\sigma_1 = 1 < \sigma_2 = \sigma_q = 1.5$.




(a) Optimality gap vs. $\tau_1$ (b) Optimality gap vs. disp($\boldsymbol{\tau}$) (c) $\mathbb{E}$[Optimality gap] vs. disp($\boldsymbol{\tau}$)

Figure 2: When does $\tau$-weighting dominate?

Figure 2(a) shows the optimality gaps under each mechanism as a function of player 1's token holdings $\tau_1$, and illustrates the dominance set $\mathcal{L}$, which is represented by a simple interval $\tau_1 \in [1/2, \tau^c]$, with $\tau^c = \frac{3}{2} - \frac{2\sigma_1^2}{\sigma_1^2+\sigma_2^2}$ and $\tau_2 = 1 - \tau_1$. We highlight two main takeaways. First, the basic intuition that $\tau$-weighting dominates when the better-informed player has more tokens does not always hold. If this player has too many tokens ($\tau_1 \geq \tau^c \sim 0.9$), $1/n$-weighting outperforms. Second, $\mathcal{G}_\tau$ has a minimum (reached at $\tau_1 \sim 0.7$) which represents the "best possible" allocation of tokens, and at this point $\tau$-weighting, is understandably better. However, as the platform does not control token allocation, it is more meaningful to consider the entire range of possible realizations of $\tau_1$ in the interval $[0, 1]$. Doing so reveals that $\tau$-weighting dominates only around 40% of the time. More importantly, when it does dominate, it tends to do so by much lower margin than in cases where it lags behind (e.g., compare $\mathcal{G}_\tau$ to $\mathcal{G}_{1/n}$ in Figure 2(a), as $\tau_1 \to 0.7$, and then as $\tau_1 \to 0$), suggesting that $1/n$-weighting might be the better mechanism "on average."

Figure 2(b) by-and-large confirms this result, but nuances it further by examining how the optimality gap changes with disp($\boldsymbol{\tau}$). To construct the values of disp($\boldsymbol{\tau}$), we simulate independent realizations of vector $\boldsymbol{\tau} = \{\tau_1, \tau_2\}$ drawn from a uniform $[0, 1]$ (normalized so that $\tau_1 + \tau_2 = 1$). The results show that the relationship between $\mathcal{G}_\tau$ and disp($\tau$) is not trivial, and is represented by two "branches" (reflecting the fact that disp($\boldsymbol{\tau}$) is symmetric under permutation of the players, *i.e.,* disp(($\tau_1, \tau_2$) = disp($\tau_2, \tau_1$)). The average value of $\mathcal{G}_\tau \sim 0.15 > \mathcal{G}_{1/n} \sim 0.10$, and hence, the optimality gap under $1/n$-weighting is $\sim 1/3$ lower on average. Finally, Figure 2(c) digs one step deeper by showing how the expectation over $\mathcal{G}_\tau$, $\mathbb{E}[\mathcal{G}_\tau]$, increases with disp($\boldsymbol{\tau}$) (this is roughly equivalent to averaging the two branches from Figure 2(b)).[11]

These results, as well as the extended analysis in Appendix A.4, support the following conclusion.

**Numerical Result 1.** *When players have heterogeneous token holdings and signal variances, then on average, as token dispersion increases, i) $\mathcal{G}_\tau$ increases and ii) $1/n$-weighting increasingly outperforms $\tau$-weighting.*

To summarize, the results suggest that when all players follow truthful voting strategies, $1/n$ weighting outperforms $\tau$-weighting, on average (across a uniform spectrum of possible token holding realizations). We postface these results with the following caveat: the entire analysis in Section 3 has so far ignored possible endogeneity in acquiring information. Presumably, strategic players may have incentive to exert effort to improve their information, and this could plausibly affect these insights. Section 4 is devoted to analyzing these endogenous information acquisition effects.

# 4 Crowdsourcing Effort

Having analyzed player voting behavior in Section 3, we now turn our attention to an important feature these crowdsourcing platforms were designed around: the incentives that players have to endogenously *improve* the platform's accuracy. Marketing materials often claim token weighting provides proper effort-exerting incentives. For instance, in the case of TCR's:

> "Token-curated registries are decentrally-curated lists with intrinsic economic incentives for token holders to curate the list's contents judiciously." — Goldin (2017).

This section examines this claim in depth. To ease exposition, we focus on effort-exerting incentives when all players are strategic. As a reminder, in this case we showed that first best variance is fully restored in the exogenous setting of Section 3. As we shall see, even in this "best case" scenario, first-best platform variance can no longer be restored in the endogenous setting. The main implications of the analysis extend to the more complicated situations in which the platform is composed of a mixture of player types (and would be more pronounced in these cases).

For the rest of the section, it will be more convenient to work with signal precisions as opposed to signal variances, $\rho_i \stackrel{\text{def}}{=} \frac{1}{\sigma_i^2}, i \in \{1, \ldots, n\}$.

[11] To be more precise, for Figure 2(c), we ran a simulation drawing 50,000 independent realizations of $\boldsymbol{\tau}$. To calculate $\mathbb{E}[\mathcal{G}_\tau]$, we discretized disp($\tau$) into 50 bins of equal length, recording the mean $\mathcal{G}_\tau$ within each bin. To preserve statistical significance, we discarded bins with less than 50 total values.

## 4.1 Model with Endogenous Information Acquisition

We begin by expanding the model in Section 2 to account for endogenous information acquisition. To this end, we introduce an additional stage to the game: prior to observing his signal and choosing his vote, player $i$ has the option to exert effort, $u_i$, to improve the precision of his signal from $\rho_i$ to $\rho_i g(u_i)$, at a cost of $c(u_i)$. We refer to $g(\cdot)$ and $c(\cdot)$ as the precision improvement and effort cost functions, respectively. This new setting is a two-stage game similar in spirit to Colombo et al. (2014), that is, in stage one, players simultaneously choose how much effort to exert, and in stage two, players simultaneously choose their votes after observing their private signals. We seek to characterize the PBNE in linear voting strategies, of this now two-stage game.

Formally, this extension renders the private signal in (1) a function of one's effort,

$$s_i(u_i) = q + \epsilon_i(u_i), \tag{7}$$

where $\epsilon_i(u_i)$ is normally distributed, with $\mathbb{E}[\epsilon_i(u_i)] = 0$ and $\text{Var}[\epsilon_i(u_i)] = \frac{1}{\rho_i g(u_i)}$. To ensure that functions $c(\cdot)$ and $g(\cdot)$ have desirable properties, e.g., effort should lead to improved information, we make the following basic assumptions.

**Assumption 2.** *Precision improvement $g(u)$ and effort cost $c(u)$ are continuous differentiable functions with $g(u) \geq 0, c(u) \geq 0, \forall u \geq 0$ (positive), $g'(u) \geq 0, c'(u) \geq 0, \forall u \geq 0$ (increasing), and $g(0) = 1$, $c(0) = 0$ (boundary conditions).*

The boundary conditions are set such that if zero effort is exerted in stage one, the outcome of the game in stage two is identical to that of Section 3 (with strategic agents).

As before, votes are linear in the signal, which means that players' voting strategy is determined by the weight, $\alpha_i$, they place on their signal, but votes now also depend on effort levels since they affect the signal, $v_i = \alpha_i s_i(u_i) + (1-\alpha_i)\mu$. Therefore, the platform's aggregate quality estimate $\hat{q} = \sum_{i=1}^n \tau_i v_i$ now also depends on effort. We use the notation $\hat{q}(\boldsymbol{u}, \boldsymbol{v})$ with $\boldsymbol{u} = \{u_1, \ldots, u_n\}$ and $\boldsymbol{v} = \{v_1, \ldots, v_n\}$, to make the dependence explicit when necessary.

Up until this stage, we have kept the platform payoff function as general as possible. We now specialize it to a standard form satisfying Assumption 1. In particular, the platform's payoff is quadratic in the aggregate error $q - \hat{q}$, that is, $\pi = k_1 - k_2(q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}))^2$, where $k_1 \geq 0, k_2 > 0$ are constants.[12] Specializing the payoff enables us to keep functions $g(\cdot)$ and $c(\cdot)$ as general as possible and shift the focus of the analysis on understanding their effects on player effort incentives.

[12]The constants $k_1, k_2$ are added for additional flexibility and to keep the value positive (if $k_1$ is large enough), but alternatively, $k_1$ could be set to zero, $k_2$ set to one, and $\pi$ could be interpreted as relative utility, rather than value.

In Section 3, player $i$ maximized his payoff over his vote $v_i$. Here, player $i$ maximizes over both effort level and vote:

$$\max_{u_i \geq 0} \left\{ -c(u_i) + \tau_i \max_{v_i \in \mathbb{L}} \mathbb{E}\left[ k_1 - k_2(q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}))^2 \;\middle|\; s_i \right] \right\}, \tag{8}$$

where the inner maximization corresponds to the second stage in which players vote after observing their signals, and the outer maximization corresponds to the first stage in which players choose their effort levels.

Note, although the payoff function is quadratic in the error, the problem is not quadratic in all decisions, which potentially makes the optimization non-trivial. To proceed, we first use existing results from Section 3 to solve the second stage. This allow us to express the players' second-stage optimal voting strategies as a function of effort levels, which reduces the two-stage optimization problem (8) to a simpler single-stage problem optimizing over effort levels.

**Lemma 2** (Problem Reduction)**.**

*(i) At the second stage of the game, given a fixed vector* $\boldsymbol{u}$*, player* $i$*'s optimal vote is* $v_i^*(\boldsymbol{u}) = \alpha_i^*(\boldsymbol{u})s_i + (1 - \alpha_i^*(\boldsymbol{u}))\mu$*, where* $\alpha_i^*(\boldsymbol{u})$ *and the platform's resulting variance are given by:*

$$\alpha_i^*(\boldsymbol{u}) = \frac{\rho_i g(u_i)}{\tau_i \left( \rho_q + \sum_{j=1}^n \rho_j g(u_j) \right)}, \quad \textit{and} \quad V(\boldsymbol{u}) = \left( \rho_q + \sum_{i=1}^n \rho_i g(u_i) \right)^{-1}. \tag{9}$$

*(ii) At the first stage of the game, player* $i$*'s profit maximization problem* (8) *is equivalent to the following cost minimization problem over effort level:*

$$\min_{u_i \geq 0} \mathcal{C}_i(\boldsymbol{u}) = \tau_i V(\boldsymbol{u}) + c(u_i). \tag{10}$$

For convenience, we define the platform's precision $\rho(\boldsymbol{u}) \stackrel{\text{def}}{=} V(\boldsymbol{u})^{-1} = \rho_q + \sum_{i=1}^n \rho_i g(u_i)$.

## 4.2 Equilibrium

Before stating the equilibrium, we discuss the conditions under which one exists. From Lemma 2, it suffices to focus on problem (10). When $\mathcal{C}_i(\boldsymbol{u})$ is continuous and convex with respect to $u_i$ for $u_i \geq 0$, there exists at least one equilibrium (Rosen 1965). Given $\mathcal{C}_i(\cdot)$ is the sum of two functions, it is convex if both functions are convex. While it is generally common to assume convex cost of information acquisition $c(\cdot)$ (see, e.g., Vives 2011), convexity of $\tau_i V(\boldsymbol{u})$ depends on the properties of the information improvement function $g(u_i)$. We have the following result.

**Lemma 3** (Convexity of $V(\cdot)$)**.** *$V(\boldsymbol{u})$ is convex in $u_i$ iff*

$$\frac{\rho_i}{\rho(\boldsymbol{u})} \geq \frac{1}{2}\frac{g''(u_i)}{g'(u_i)^2}. \tag{11}$$

*A sufficient condition is $g(\cdot)$ linear or concave.*

Lemma 3 has important implications. First, note that $\rho_i \geq 0$ and $\rho(\boldsymbol{u}) \geq 0$ by construction, thus the left-hand side is positive. It follows that convexity of $V(\cdot)$ trivially holds when $g(\cdot)$ is linear as $g''(\cdot) = 0$ in this case. For instance, $g(u_i) = 1 + u_i$, which also satisfies Assumption 2 ($g > 0, g' > 0$ and $g(0) = 1$). Convexity of $V(\cdot)$ also trivially holds when $g(\cdot)$ is concave as $g''(\cdot) \leq 0$ in this case.

Up until now, we have been working with general functions $g(\cdot)$ and $c(\cdot)$, and our results, while also general, have not addressed the equilibrium outcome. To make headway, we need to impose some additional structural requirements on $c(\cdot)$ and $g(\cdot)$. To see why, consider the first-order conditions of (10) $\tau_i V' + c' = 0, \forall i$. Using the expression for $V'$ (see equation (33) in the appendix), these are equivalent to the following system of non-linear equations

$$\begin{aligned}
\tau_1\rho_1 &= \left(\frac{\partial c(u_1)}{\partial u_1}\bigg/\frac{\partial g(u_1)}{\partial u_1}\right)\rho(u_1,\ldots,u_n)^2 \\
&\vdots \\
\tau_n\rho_n &= \left(\frac{\partial c(u_n)}{\partial u_n}\bigg/\frac{\partial g(u_n)}{\partial u_n}\right)\rho(u_1,\ldots,u_n)^2.
\end{aligned} \tag{12}$$

While such non-linear systems are generally intractible, the structure of these equations motivates the following additional assumption that will facilitate the computation of the equilibrium.

**Assumption 3.** *Let $k > 0$ be a constant and suppose $c$ and $g$ satisfy the following equation:*

$$c'/g' = k \cdot g^2. \tag{13}$$

In the Appendix A.5, we show that a broad class of linear and concave forms of $g(\cdot)$ (which by Lemma 3, suffice for equilibrium existence), and corresponding functions $c$, satisfy this requirement. For instance, consider $g(u_i) = 1 + u_i$, thus $g'(u_i) = 1$. Condition (13) (with $k = 1$) implies $c'(u_i) = (1 + u_i)^2$, and thus, $c(u_i) = (1/3)(1 + u_i)^3$, which is convex, positive increasing for $u_i \geq 0$.

Leveraging (13) and some additional properties, we obtain that the system (12) not only can have a unique solution, but one that is also characterizable in closed form.

**Theorem 1** (Equilibrium)**.** *Let vectors* $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_n\}$*,* $\boldsymbol{\rho} = \{\rho_1, \ldots, \rho_n\}$*. Suppose* $c(\cdot)$ *and* $g(\cdot)$ *satisfy Assumptions 2 and 3,* $c(\cdot)$ *is convex, and* $g(\cdot)$ *is either linear, or a concave bijective (invertible) function. Then, the effort-exerting game has a unique equilibrium, characterized by votes* $v_i^*(\boldsymbol{u}^*) = \alpha_i^*(\boldsymbol{u}^*)s_i + (1-\alpha_i^*(\boldsymbol{u}^*))\mu$ *where the signal weights* $\alpha_i^*(\boldsymbol{u}^*)$ *are given in* (9)*, and effort levels* $u_i^*$ *are given by*

$$u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}) = g^{-1}\Big(f_i^*(\boldsymbol{\tau}, \boldsymbol{\rho})\Big), \text{ with } f_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} \sum_{j=1}^{n} \left(\tau_j \rho_j^3\right)^{1/2}\right)^{1/2} - \rho_q}{2 \frac{1}{(\rho_i \tau_i)^{1/2}} \sum_{j=1}^{n} \left(\tau_j \rho_j^3\right)^{1/2}}. \tag{14}$$

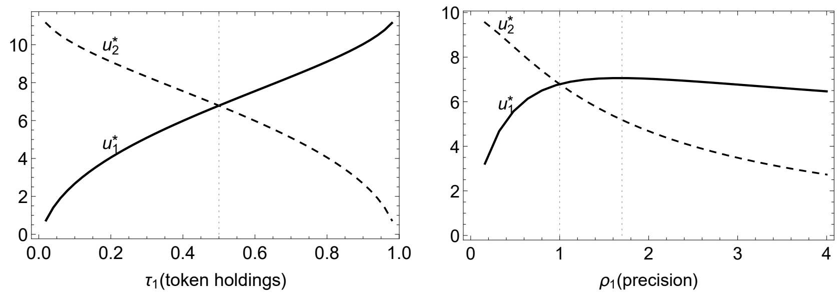
Having derived the equilibrium, we now examine how equilibrium effort levels depend on token holdings and information precision.

**Proposition 5** (Comparative Statics)**.** *Suppose the conditions of Theorem 1 are satisfied, then equilibrium effort levels (i) monotonically increase in token holdings and (ii) are non-monotonic in information precisions. Further, consider two hypothetical players* $i$ *and* $i'$ *and suppose they have the same precision, then (iii)* $\tau_i > \tau_{i'} \Rightarrow u_i^* > u_{i'}^*$*; suppose instead the two players have the same token holdings, then (iv)* $\rho_i > \rho_{i'} \Rightarrow u_i^* > u_{i'}^*$*.*

Proposition 5 (i) states that the more tokens a player has, the more effort he provides in equilibrium – a desirable feature often touted in marketing materials. Relatedly, part (iii) can be interpreted as an equivalent statement on fairness – when comparing two equally informed players, the one with more tokens always exerts more effort in equilibrium. However, as we shall show in Section 4.3, these desirable features are not attributable to the $\tau$-weighted aggregation mechanism itself.

Part (ii) highlights that there is a more subtle (non-monotonic) relationship between the amount of effort a player exerts and his information precision. Part (iv) shows that when comparing two players with equal token holdings, the one with better information precision exerts more effort in equilibrium. In other words, there is some amount of "free-riding" from lesser-informed players, which is not necessarily a desirable feature.

To illustrate these results, Figure 3 shows equilibrium effort levels in a $n = 2$-player game assuming a linear information improvement function $g(x) = 1 + \eta x$, with $\eta > 0$. Figure 3(a) (parameters $\eta = \rho_1 = \rho_2 = \rho_q = 1$) highlights the results of Proposition 5 parts (i) and (iii). That is, player one's effort increases with his share of the tokens $\tau_1$, and player one exerts more effort than player two when $\tau_1 > 1/2$. Figure 3(b) (parameters $\eta = \tau_1 = \tau_2 = 1/2, \rho_2 = \rho_q = 1$) shows the more subtle relationship between effort levels and information precision described in Proposition 5 parts (ii) and (iv). The first vertical line marks the point where player one's precision and effort exceed player two's, and the second vertical line highlights the non-monotonicity.




(a) Effort $u_1^*$ and $u_2^*$ vs. tokens $\tau_1$ (b) Effort $u_1^*$ and $u_2^*$ vs. precision $\rho_1$

Figure 3: Equilibrium effort levels in a 2-player game with a linear information improvement function

For the rest of the analysis, we assume all results are subject to the conditions of Theorem 1 and we will work with the linear form of $g(\cdot)$ described above and in Appendix A.5.1.

## 4.3 Centralized Solution and Optimality Gap

Having derived player equilibrium effort levels, we now compare the resulting platform accuracy they obtain to that of a centralized setting.

In the centralized case, we assume the platform is maximizing the sum of all expected player payoffs. Noting that $\sum_{i=1}^{n} \tau_i V(\boldsymbol{u}) = V(\boldsymbol{u}) \sum_{i=1}^{n} \tau_i = V(\boldsymbol{u})$, and using the equivalence between maximizing payoffs and minimizing variance from Lemma 1, the centralized problem can be written as a minimization problem over the platform variance and effort costs given by

$$\min_{\boldsymbol{u} \geq 0} V(\boldsymbol{u}) + \sum_{i=1}^{n} c(u_i). \tag{15}$$

Convexity of $V(\cdot)$ (see Lemma 3) and $c(\cdot)$ suffice to guarantee a unique solution. Denote by $\boldsymbol{u}^{fb}$ the vector representing this solution and let $V^{fb} \stackrel{\text{def}}{=} V(\boldsymbol{u}^{fb})$. Let $V^{eq} \stackrel{\text{def}}{=} V(\boldsymbol{u}^*)$, with $\boldsymbol{u}^* = \{u_1^*(\boldsymbol{\tau}, \boldsymbol{\rho}), \ldots, u_n^*(\boldsymbol{\tau}, \boldsymbol{\rho})\}$ given by (14), be the equilibrium platform variance of the decentralized game, and define the optimality gap between the two as

$$\mathcal{G} = V^{eq} - V^{fb}. \tag{16}$$

**Proposition 6.** *Suppose the assumptions from Theorem 1 hold, and let* $\mathbf{1} = \{1, \ldots, 1\}$ *be an n-dimensional vector of ones, then*
*(i) the optimal effort levels of the centralized platform are given by* $u_i^{fb}(\boldsymbol{\rho}) = u_i^*(\mathbf{1}, \boldsymbol{\rho}), \forall i \in \{1, \ldots, n\}$*;*
*(ii) the effort levels of the centralized platform exceed those of the decentralized one,* $\boldsymbol{u}^{fb} > \boldsymbol{u}^*$*;*
*(iii) the optimality gap is positive,* $\mathcal{G} > 0$*.*

Part (i) of Proposition 6 shows that the effort levels of the centralized platform have a simple connection to the equilibrium effort levels of the decentralized platform in (14). Mathematically, they can be obtained by setting $\tau_i = 1, \forall i$. This implies the distribution of token holdings is irrelevant from the centralized platform's perspective; intuitively, when it comes to assigning effort levels to players, only the vector of information accuracy, $\boldsymbol{\rho}$, matters. Part (ii) follows from Proposition 5(i) which implies $u_i^*(\mathbf{1}, \boldsymbol{\rho}) > u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}), \forall n > 1$, as long as $\tau_i \neq 1, \forall i$, that is, no single player has all the tokens. Part (iii) follows from part (ii) and the fact that $V(\boldsymbol{u})$ is decreasing in $\boldsymbol{u}$ (see the proof of Proposition 6).

Unlike our results in Section 3, where strategic players can fully restore first best, here, the optimality gap is strictly positive. This is because in contrast to voting, providing effort is costly, and each player stands to benefit differently depending on his token holdings, which implies that player $i$'s individual profit incentives are no longer aligned with the platform's.

Having shown that first-best accuracy can no longer be restored in the game with effort, even in the "best case" scenario where all agents are strategic, we now examine how the resulting optimality gap is affected by i) the aggregation mechanism, comparing $\tau$- and $1/n$-weighted schemes, and ii) the dispersion in token holdings.

(i) **Irrelevance of Aggregation Mechanism.** Recall that in Section 3, we showed the aggregation mechanism to be irrelevant if all players are strategic. This result continues to hold in the broader game with effort. To see why, consider the voting game (8) and note that for a fixed vector $\boldsymbol{u}$, player $i$'s optimization problem in the second stage parallels the original voting game in (4) analyzed in Section 3 (the two are in fact identical when $\boldsymbol{u} = 0$). When all players are strategic, they adjust their votes in the second stage to "undo" the platform's aggregation mechanism and recover the first best accuracy achievable for a given vector $\boldsymbol{u}$. In other words, although the platform's choice of aggregation mechanism influences how individual players vote, it does not influence the final voting outcome, once all votes are aggregated. This also means that the optimality gap is not impacted by the choice of the aggregation mechanism (formal statements are provided in the proof of Lemma 2).

(ii) **Relevance of Token Dispersion.** It is important to emphasize that the previous result does *not* imply that the distribution of tokens is irrelevant. In fact, Theorem 1 shows the opposite: each player's equilibrium effort level depends on the entire vector of token holdings $\boldsymbol{\tau}$. This is because the distribution of tokens directly affects each player's expected payoff in the first stage of the game, due to the proportionality of returns. Recall individual payoffs are defined as $\tau_i\pi$, that is, players with more tokens stand to benefit more from an increase in platform value, and in turn, are incentivized to exert more effort. Ironically, this runs contrary to some of the claims made by $\tau$-weighted platform operators, regarding the role played by the aggregation mechanism in providing adequate effort-exerting incentives. The fact that this effect is tied to the proportionality of returns means that it would persist even if platforms were to adopt an equally weighted $1/n$ mechanism. This suggests it is particularly meaningful to explore question ii) regarding the impact of token dispersion, in some depth. We do so next.

## 4.4 The Impact of Token Dispersion

In this section, we examine the impact of token dispersion on the optimality gap. Before proceeding, we summarize the equilibrium player actions and derive the platform variances in the decentralized and centralized (first-best) settings in Table 2. Within the decentralized setting, it will be useful to distinguish between a special homogeneous case in which all players hold $1/n$ tokens, and the general heterogeneous case in which token holdings are dispersed. Abusing notation for simplicity, let $\boldsymbol{n}^{-1} \stackrel{\text{def}}{=} \{1/n, \ldots, 1/n\}$ and denote by $V_{1/n}^{eq} \stackrel{\text{def}}{=} V^{eq}(\boldsymbol{\tau} \to \boldsymbol{n}^{-1})$, the platform variance in the purely homogeneous case. This type of distinction is not necessary in the centralized case given token holdings are irrelevant in that setting. We emphasize this notation is fundamentally different than the one adopted in Section 3, where the subscript $1/n$ represented the platform choosing an equally weighted aggregation mechanism. Here, subscript $1/n$ represents the special case of homogeneous holdings under $\tau$-weighted aggregation. Without loss of generality, we display all expressions taking the limit $\sigma_q \to \infty$ to ease exposition.
We proceed by first examining the simple case in which all players have the same precision, and then analyze the more complicated heterogeneous case.

### Optimality gap with homogeneous precisions

Similarly to (16), let $\mathcal{G}_{1/n} \stackrel{\text{def}}{=} V_{1/n}^{eq} - V^{fb}$ denote the optimality gap between the equilibrium platform variance when all players hold $1/n$ tokens, and first best.

Table 2: Equilibrium player actions and platform variance (as $\sigma_q \to \infty$).

| **Setting** | **Equilibrium Votes and Effort** | **Platform Variance** |
|---|---|---|
| Decentralized, dispersed holdings | $v_i(\alpha_i^*(u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}))); u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho})$ | $V^{eq} = k^{1/4} \left( \sum_{i=1}^n \rho_i^{3/2} \sqrt{\tau_i} \right)^{-1/2}$ |
| Decentralized, $1/n$ holdings | $v_i(\alpha_i^*(n^{-1}, u_i^*(\boldsymbol{n}^{-1}, \boldsymbol{\rho}))); u_i^*(\boldsymbol{n}^{-1}, \boldsymbol{\rho})$ | $V_{1/n}^{eq} = k^{1/4} \left( (1/\sqrt{n}) \sum_{i=1}^n \rho_i^{3/2} \right)^{-1/2}$ |
| Centralized (first best) | $v_i(\alpha_i^*(1, u_i^*(\boldsymbol{1}, \boldsymbol{\rho}))); u_i^*(\boldsymbol{1}, \boldsymbol{\rho})$ | $V^{fb} = k^{1/4} \left( \sum_{i=1}^n \rho_i^{3/2} \right)^{-1/2}$ |

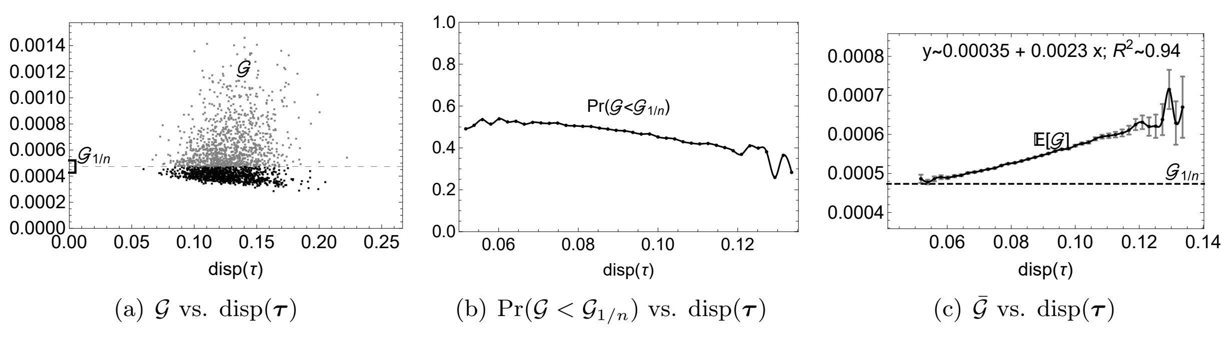
**Proposition 7.** *Under the conditions of Theorem 1, when all players strategically choose effort levels and votes, and have homogeneous precisions, i.e.,* $\rho_i = \rho$, $\forall i \in \{0, \ldots, n\}$, *we have* $\mathcal{G} \geq \mathcal{G}_{1/n} > 0$, *that is, the platform would prefer an even allocation of tokens.*

It is interesting to note that a homogeneous allocation is preferred despite the fact that players with more tokens exert more effort in equilibrium (see Proposition 5). This is (in part) due to the convexity of $V^{eq}$ with respect to $\boldsymbol{\tau}$ (see the proof of Proposition 7 for a step-by-step derivation).

We now examine the more general case with heterogeneous precisions.

### Optimality gap with heterogeneous precisions

The set of $\boldsymbol{\tau}$'s for which a dispersed allocation dominates is given by $\{\boldsymbol{\tau} \mid V^{eq}(\boldsymbol{\tau}) \leq V_{1/n}^{eq} \cap \boldsymbol{1}'\boldsymbol{\tau} = 1\}$. We examine dominance numerically, considering a platform with $n = 20$ players, described by a fixed vector $\boldsymbol{\sigma}$ (split evenly across the interval $[0, 1]$, meaning the average signal precision is 0.5). As before, we draw realizations of vector $\boldsymbol{\tau}$ from a uniform in $[0, 1]$ (normalized so that the components sum to 1). For each one of those draws, we record the corresponding value of $\mathcal{G}$ and disp($\boldsymbol{\tau}$).




(a) $\mathcal{G}$ vs. disp($\boldsymbol{\tau}$) (b) $\Pr(\mathcal{G} < \mathcal{G}_{1/n})$ vs. disp($\boldsymbol{\tau}$) (c) $\bar{\mathcal{G}}$ vs. disp($\boldsymbol{\tau}$)

Figure 4: Optimality gap $\mathcal{G}$ with heterogeneous precisions

Figure 4(a) shows the optimality gap $\mathcal{G}$ as a function of disp($\boldsymbol{\tau}$). The single point represented by an empty rectangle at the $x$-axis origin is the optimality gap resulting from an even allocation of tokens, $\mathcal{G}_{1/n}$. By definition, $\text{disp}(\boldsymbol{n}^{-1}) = 0$ (there is no dispersion when all players have the same number of tokens). From this point, we trace the horizontal dashed line. Points below this horizontal line, in black, represent dominance of dispersed token allocations, over the $1/n$ allocation. Points above this line, in gray, represent the reverse. There is a positive, albeit low correlation (~10%) between $\mathcal{G}$ and $\text{disp}(\boldsymbol{\tau})$, suggesting that, on average, the optimality gap slightly increases with $\text{disp}(\boldsymbol{\tau})$, though the effect is weak.

To bring to light the intricacies of their relationship, we run a more extensive simulation: Figure 4(b) shows that the probability that dispersed tokens are preferred, $\Pr(\mathcal{G} < \mathcal{G}_{1/n})$, tends to decrease with $\text{disp}(\boldsymbol{\tau})$ (going from $\sim 50\%$ to $\sim 20\%$). Figure 4(c) expands the analysis beyond probabilities to take into account magnitudes, and shows that the average $\mathbb{E}[\mathcal{G}]$ tends to increase with $\text{disp}(\boldsymbol{\tau})$.[13] These results support the following conclusion.

**Numerical Result 2.** *When all players strategically choose effort levels and votes, and have heterogeneous precisions, equilibrium effort levels decrease and the optimality gap increases, on average, with token dispersion.*

## 4.5 Optimal Token Allocation

Though the previous analysis suggests it is "on average" preferable to limit token dispersion across players, it does not account for the distribution of precisions which can be used to better inform the platform on the preferred allocation of tokens. Figure 4(b) shows that between 20% to 50% of the time, the platform would be better off with dispersed tokens. It is therefore of interest to examine these instances further.

Consider the equilibrium platform variance $V^{eq}(\boldsymbol{\tau})$, where we are now being explicit about the dependence on the vector of tokens ($\boldsymbol{\tau}$). The allocation of tokens which minimizes the platform's variance with strategic players, is given by $\boldsymbol{\tau}^* \stackrel{\text{def}}{=} \{\text{argmin}_{\boldsymbol{\tau}}\, V^{eq}(\boldsymbol{\tau}),\ \text{subject to } \mathbf{1}'\boldsymbol{\tau} = 1\}$. For the linear effort model and taking the $\lim_{\sigma_q \to \infty} V^{eq}(\boldsymbol{\tau})$ to simplify expressions, the optimal allocation is obtainable in closed form:

$$\tau_i^*(\boldsymbol{\rho}) = \left(1 + \frac{1}{\rho_i^3} \sum_{j \neq i} \rho_j^3\right)^{-1}. \tag{17}$$

As can be intuitively expected, player $i$'s optimal token holdings are a function of all players'

[13] To be more precise, in both cases, we ran a simulation drawing 1mm independent realizations of $\boldsymbol{\tau}$ from a standard (normalized) uniform distribution. We then discretized $\text{disp}(\tau)$ into 50 bins of equal length, recording $\Pr(\mathcal{G} < \mathcal{G}_{1/n})$ and $\mathbb{E}[\mathcal{G}]$ within each bin. To preserve statistical significance, we discarded bins with less than 50 total values.

precisions: they increase in his own precision, and decrease in others' precisions.

Next, consider the Euclidean distance $d(\boldsymbol{\tau}, \boldsymbol{\tau}^*)$, which measures how "far apart" a random vector $\boldsymbol{\tau}$ is from the optimal allocation $\boldsymbol{\tau}^*$. Using the same simulation parameters as in Section 4.4, Figure 5 shows that the optimality gap tends to increase with this distance. Furthermore, there is a positive correlation of 85% between the two, suggesting a strong effect. This leads to our last numerical result.

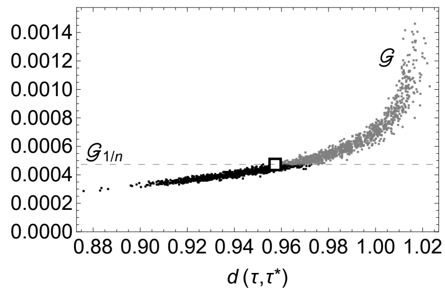



Figure 5: $\mathcal{G}$ versus $d(\boldsymbol{\tau}, \boldsymbol{\tau}^*)$.

**Numerical Result 3.** *When all players strategically choose effort levels and votes, and have heterogeneous precisions, i) the optimality gap increases, on average, with* $d(\boldsymbol{\tau}, \boldsymbol{\tau}^*)$*; ii) there exists a critical distance above which all even allocations dominate dispersed allocations, and below which, all dispersed allocations dominate even allocations.*

Note, the numerical results are generally robust when changing: the number of players $n$, the fixed vector $\boldsymbol{\rho}$, and the metric chosen to compare vectors (e.g., similar results are obtainable if using the geometric angle between vectors, as opposed to the Euclidean distance).

# 5 Conclusion & Limitations

Many blockchain-based platforms have deployed voting mechanisms that use some form of token-weighted aggregation. These platforms rely on the argument that token weighting incentivizes users to provide higher-quality votes, which in turn improves the overall accuracy of the platform. Our results show that this intuition is at best only partially correct. In many cases, the platform could achieve equal, or sometimes better accuracy (higher overall payoffs) by pursuing a different aggregation strategy and/or by limiting token dispersion across users.

Like all stylized models, some of our assumptions place limitations on the scope of the paper. For instance, our model assumes a continuous voting and outcome space in $\mathbb{R}$. In some cases, votes are constrained to be binary, i.e., users can vote content up or down/like or dislike, and in others, votes are constrained to be in discrete ranges, e.g., Uber ratings following a completed trip allow users to provide only discrete feedback levels, 1-5 stars. Constraining the vote space would be an interesting direction to explore.

Second, some of our results rely on players knowing the token holdings, signal precisions, and/or the strategy types of other players. Relaxing these types of assumptions is challenging, but could lead to interesting insights, such as whether platforms should purposely obfuscate information or disseminate it as much as possible.

As mentioned in the introduction, token-weighted voting has been adopted for many different applications, and our model paints these systems with a broad enough brush to capture some of the common economic tradeoffs they all share. As such, this paper should be viewed as a first step in expanding our theoretical understanding of these systems. These limitations could represent interesting directions for future work in this space.

## References


Acemoglu, D., K. Bimpikis, A. Ozdaglar. 2014. Dynamics of information exchange in endogenous social networks. *Theoretical Economics* **9**(1) 41–97.

Acemoglu, D., M. A. Dahleh, I. Lobel, A. Ozdaglar. 2011. Bayesian learning in social networks. *The Review of Economic Studies* **78**(4) 1201–1236.

Alaei, S., A. Malekian, M. Mostagir. 2016. A dynamic model of crowdfunding. Working Paper.

Araman, V. F., R. Caldentey. 2016. Crowdvoting the timing of new product introduction. *Available at SSRN 2723515* .

Arcange. 2019. the daily analytical report about the steem blockchain. https://steemit.com/statistics/@arcange/steem-statistics-20190327-en. Accessed 2019-04-22.

Asgaonkar, A., B. Krishnamachari. 2018. Token curated registries-a game theoretic approach. *arXiv preprint arXiv:1809.01756* .

Ashu, D. Lv. 2018. Carbon vote. http://carbonvote.com/. Accessed 2018-09-28.

Ashu, D. Lv. 2019. Vote: TheDAO hard fork. http://v1.carbonvote.com/. Accessed 2019-04-22.

Babich, V., G. Hilary. 2018. What operations management researchers should know about blockchain technology. Georgetown University Working Paper.

Babich, V., S. Marinesi, G. Tsoukalas. 2019. Does crowdfunding benefit entrepreneurs and venture capital investors? *M&SOM, Forthcoming* .

Banzhaf III, J. F. 1964. Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev.* **19** 317.

Belavina, E., S. Marinesi, G. Tsoukalas. 2019. Designing crowdfunding platform rules to deter misconduct. *Management Science, Forthcoming* .

Bergemann, D., J. Välimäki. 1997. Market diffusion with two-sided learning. *The RAND Journal of Economics* 773–795.

Bhatia, A., R. Giometti, A. Nicolas. 2018. Sapien: Decentralized social news platform. https://www.sapien.network/static/pdf/SPNv1_3.pdf. Accessed 2018-09-28.

Biais, B., C. Bisiere, M. Bouvard, C. Casamatta. 2017. The blockchain folk theorem. Toulouse School of Economics, TSE-817 Working Paper.

Biais, B., C. Bisiere, M. Bouvard, C. Casamatta, A. J. Menkveld. 2018. Equilibrium bitcoin pricing. *Available at SSRN* .

Bimpikis, K., S. Ehsani, M. Mostagir. 2015. Designing dynamic contests. *EC*. 281–282.

Cao, S., L. W. Cong, B. Yang. 2018. Auditing and blockchains: Pricing, misstatements, and regulation. *Misstatements, and Regulation (Oct 9, 2018)* .

Catalini, C., J. Gans. 2017. Some simple economics of the blockchain. MIT Sloan Research Paper No. 5191-16.

Chandrasekhar, A. G., H. Larreguy, J. P. Xandri. 2015. Testing models of social learning on networks: Evidence from a lab experiment in the field. Tech. rep., National Bureau of Economic Research.

Chod, J., E. Lyandres. 2018. A theory of icos: Diversification, agency, and information asymmetry. Boston College Working Paper.

Chod, J., N. Trichakis, G. Tsoukalas. 2018. Blockchain and the value of operational transparency for supply chain finance. *Working Paper* .

Christoffersen, S. E., C. C. Geczy, D. K. Musto, A. V. Reed. 2007. Vote trading and information aggregation. *The Journal of Finance* **62**(6) 2897–2929.

Colombo, L., G. Femminis, A. Pavan. 2014. Information acquisition and welfare. *The Review of Economic Studies* **81**(4) 1438–1483.

Cong, L. W., Y. Li, N. Wang. 2018. Tokenomics: Dynamic adoption and valuation. *Working Paper* .

Elkind, E., G. Chalkiadakis, N. R. Jennings. 2008. Coalition structures in weighted voting games. *ECAI*, vol. 8. 393–397.

Eufemio, A. C., K. C. Chng, S. Djie. 2018. Digix's whitepaper: The gold standard in crypto assets. https://digix.global/whitepaper.pdf. Accessed 2018-09-28.

Even, Y., A. Tahbaz-Salehi, X. Vives. 2019. Information aggregation in heterogenous markets. *Working Paper* .

Gan, J., G. Tsoukalas, S. Netessine. 2019. Inventory, speculators, and initial coin offerings. *Working Paper* .

Gifford, D. K. 1979. Weighted voting for replicated data. *Proceedings of the seventh ACM symposium on Operating systems principles*. ACM, 150–162.

Goldin, M. 2017. Token-curated registries 1.0. *Online* URL https://docs.google.com/document/d/1BWWC__-Kmso9b7yCI_R7ysoGFIT9D_sfjH3axQsmB6E/.

Goldin, M., A. Soleimani, J. Young. 2017. The AdChain registry. https://adtoken.com/uploads/white-paper.pdf. Accessed 2018-09-28.

Goodman, L. 2014. Tezos-a self-amending crypto-ledger white paper. Tech. rep., The Tezos Foundation.

Hentschker, L. 2018. Decert. http://decert.io/.

Hinzen, F. J., K. John, F. Saleh. 2019. Proof-of-work's limited adoption problem. *NYU Stern School of Business* .

Iles, M. 2018. The Civil white paper. https://blog.joincivil.com/the-civil-white-paper-3e6c6f72dd9e. Accessed 2018-09-28.

Jentzsch, C. 2016. The history of the dao and lessons learned. https://blog.slock.it/the-history-of-the-dao-and-lessons-learned-d06740f8cfa5. Accessed 2018-09-28.

Kim, C. 2019. Welcome to athens: Tezos completes 'historic' first blockchain vote. https://www.coindesk.com/welcome-to-athens-tezos-completes-historic-first-blockchain-vote.

Levit, D., N. Malenko. 2011. Nonbinding voting for shareholder proposals. *The Journal of Finance* **66**(5) 1579–1614.

Lockyer, M. 2018. Token curated registry (TCR) design patterns. https://hackernoon.com/token-curated-registry-tcr-design-patterns-4de6d18efa15. Accessed 2018-09-28.

Myatt, D. P., C. Wallace. 2012. Endogenous information acquisition in coordination games. *The Review of Economic Studies* **79**(1) 340–374.

Nordmann, L., H. Pham. 1999. Weighted voting systems. *IEEE Transactions on Reliability* **48**(1) 42–49.

Ostrovsky, M. 2012. Information aggregation in dynamic markets with strategic traders. *Econometrica* **80**(6) 2595–2647.

Papanastasiou, Y., K. Bimpikis, N. Savva. 2017. Crowdsourcing exploration. *Management Science* **64**(4) 1727–1746.

Rosen, J. B. 1965. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society* 520–534.

Rosu, I., F. Saleh. 2019. Evolution of shares in a proof-of-stake cryptocurrency. *Working Paper* .

Saghafian, S., B. Tomlin, S. Biller. 2018. The internet of things and information fusion: Who talks to who? *Working Paper, Tuck School of Business No. 3124417* .

Saleh, F. 2018. Blockchain without waste: Proof-of-stake. *Working Paper* .

Shleifer, A., R. W. Vishny. 1986. Large shareholders and corporate control. *Journal of political economy* **94**(3, Part 1) 461–488.

Simon, d. l. R. 2017. Curation markets. https://docs.google.com/document/d/1VNkBjjGhcZUV9CyC0ccWYbqeOoVKT2maqX0rK3yXB20/edit. Accessed 2018-09-28.

Snyder Jr, J. M., M. M. Ting, S. Ansolabehere. 2005. Legislative bargaining under weighted voting. *American Economic Review* **95**(4) 981–1004.

Steem. 2018. Steem an incentivized, blockchain-based, public content platform. https://steem.io/steem-whitepaper.pdf. Accessed 2018-09-28.

Stouras, K. I., J. Hutchison-Krupat, R. O. Chao. 2017. The role of participation in innovation contests. *Working Papers* .

Strausz, R. 2017. A theory of crowdfunding: A mechanism design approach with demand uncertainty and moral hazard. *American Economic Review* **107**(6) 1430–76.

Taylor, A., W. Zwicker. 1993. Weighted voting, multicameral representation, and power. *Games and Economic Behavior* **5**(1) 170–181.

Terwiesch, C., Y. Xu. 2008. Innovation contests, open innovation, and multiagent problem solving. *Management science* **54**(9) 1529–1543.

Tong, Z., R. Y. Kain. 1991. Vote assignments in weighted voting mechanisms. *IEEE Transactions on Computers* **40**(5) 664–667.

Vives, X. 1988. Aggregation of information in large cournot markets. *Econometrica: Journal of the Econometric Society* 851–876.

Vives, X. 2011. Strategic supply function competition with private information. *Econometrica* **79**(6) 1919–1966.

Warren, W., A. Bandeali. 2018. 0x: An open protocol for decentralized exchange on the ethereum blockchain. https://0xproject.com/pdfs/0x_white_paper.pdf. Accessed 2018-09-28.

Yermack, D. 2017. Corporate governance and blockchains. Review of Finance Forthcoming.

# Token-Weighted Crowdsourcing

---

# Online Appendix

## A Auxiliary Results

### A.1 Examples of "Well-Behaved" and "Ill-Behaved" Payoff Functions

If there is only a single player, who receives a signal, $s$, makes a vote $v$ and receives a payoff, $\pi(q-v)$, what is the optimal vote? At first glance, it seems that the obvious strategy is to vote the conditional mean $v = \mathbb{E}[q \mid s]$. This is certainly the case if $\pi(x) = -x^2$ (see Example 1). On the other hand, this statement is actually *not* true in general.

**Example 1.** *When $\pi(x) = d - x^2$, the fact that $v = \mathbb{E}[q \mid s]$ is optimal corresponds to the fact that the mean is the best least-squares estimator.*

**Example 2** ("Price is Right" payoffs)**.** *Suppose*

$$\pi(x) = \begin{cases} \frac{1}{x^2+1} & \text{if } x < 0 \\ 0 & \text{if } x \geq 0. \end{cases}$$

*This corresponds to a "Price is Right" rule, where underestimates are valued according to their accuracy, whereas overestimates give a payoff of 0. Suppose that conditioned on the signal, $q$ has a standard normal distribution,* i.e.,

$$q|_s \sim N(0,1).$$

*In this case, setting $v = \mathbb{E}[q \mid s]$ corresponds to $v = 0$, and the expected payoff is approximately* $.175$*. On the other hand, setting $v = -\frac{1}{2}$, yields a payoff of approximately* $.213$*, thus for this payoff function, guessing the conditional mean is* not *optimal.*

**Example 3.** *Suppose that (conditioned on $s$) $q$ takes on only two values, it has "low" quality ($q = -1$) with probability one half, and "high" ($q = 1$) with probability one half. Suppose*

$$\pi(x) = \begin{cases} 1 & \text{if } |x| < 1 \\ 0 & \text{otherwise.} \end{cases}$$

*Then $\mathbb{E}[q \mid s] = 0$, but setting $v = 0$, will always yield a payoff of $0$, wheres voting $v = 1$ will yield an expected payoff of one half.*

Observe that in Example 2, the payoff function $\pi(x)$ is not symmetric, whereas in Example 3, the distribution of $q$ is not decreasing away from its mean.

### A.2 Single-Player Solution

In the single-player setting, the optimal strategy is to vote the conditional mean, *i.e.,* to set $v = \mathbb{E}[q \mid s]$. To see this, consider first that the restriction to linear strategies implies player 1's strategy is entirely determined by the weight, $\alpha$, he places on his signal. Then, since the true quality, $q$, and the player's signal are normally distributed, *i.e.,* $q \sim N(\mu, \sigma_q^2)$ and $s = q + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$, Lemma 1 shows that choosing $\alpha$ to maximize the payoff is equivalent to choosing $\alpha$ to minimize the variance $\text{Var}[q - v]$. Since $v = \alpha s + (1-\alpha)\mu$, we have

$$\begin{aligned}\text{Var}[q - v] = \text{Var}[q - \alpha s - (1-\alpha)\mu] &= \text{Var}[(1-\alpha)q - \alpha\epsilon - (1-\alpha)\mu] \\ &= \text{Var}[(1-\alpha)q] + \text{Var}[\alpha\epsilon] \\ &= (1-\alpha)^2\sigma_q^2 + \alpha^2\sigma^2 \\ &= (\sigma^2 + \sigma_q^2)\alpha^2 - 2\sigma_q^2\alpha + \sigma_q^2,\end{aligned}$$

which is minimized when $\alpha = \frac{\sigma_q^2}{\sigma_q^2+\sigma^2}$. Thus $v = \frac{\sigma_q^2}{\sigma_q^2+\sigma^2}s + \frac{\sigma^2}{\sigma_q^2+\sigma^2}\mu = \mathbb{E}[q \mid s]$. If $q$ and $s$ are *not* normally distributed, the result still holds, but the proof is more involved.

### A.3 Optimal Aggregation Mechanism with Truthful Voters

In this section, we consider the possibility of shifting the burden of optimal aggregation to the platform instead of the players. This is only relevant in the presence of truthful voters (because in their absence, strategic voters restore first best). In contrast to the centralized platform setting in which the platform receives the signals directly, here we assume the platform receives players' truthful votes $v_1, \ldots, v_n$, but cannot observe their signals. Using the equivalence between maximizing profits and minimizing variance (Lemma 1), the strategic platform's problem can be written as

$$\min_{w_i} \text{Var}\left[q - \sum_{i=1}^{n} w_i v_i\right],$$

where $v_i = \alpha_i s_i + (1-\alpha_i)\mu$ and $w_i$ is the aggregator's decision variable of how much weight to put on each incoming vote.

**Proposition 8** (Optimal Aggregation Mechanism with Truthful Voters)**.** *When players are truthful* *and submit votes $v_i(\alpha_i)$, there exists a unique optimal mapping of platform aggregation weights $w_i^*(v_i(\alpha_i)) \neq \tau_i$ that can achieve first-best accuracy, as long as $\alpha_i \neq 0$ for all $i$.*

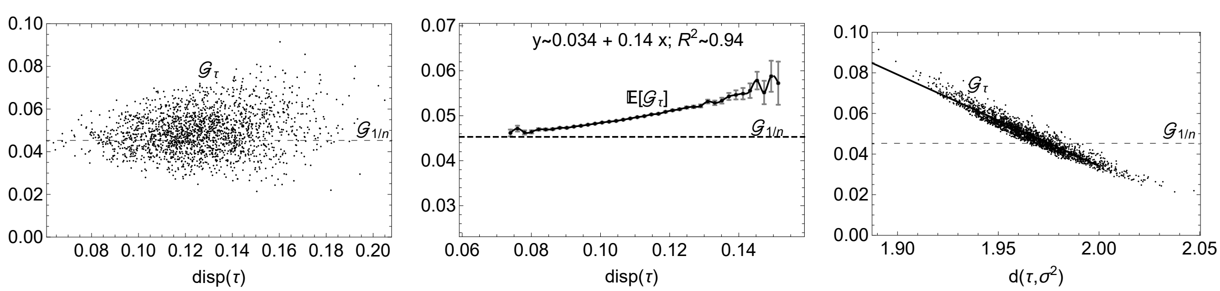
The details are provided in the proof of Proposition 8. Three important caveats are in order: (i) players must place non-zero weight on their signals, otherwise the platform cannot make a correct inference; (ii) the platform must have the same prior as the players; (iii) centralizing the aggregation mechanism in this way goes against the core philosophy these platforms were built on — that of decentralizing decisions, so may be of limited practical applicability.

Proposition 8 omits the case of strategic voters because in that case, for any nonzero set of weights $w_i$, the strategic players can adjust their $\alpha_i$ to ensure that optimality holds (as in Proposition 1).

Note, if the platform contains both truthful and strategic voters, as in Section 3.3, the platform would need to construct the aggregate vote $\hat{q} = \sum_{i=1}^n w_i v_i$ (instead of $\hat{q} = \sum_{i=1}^n \tau_i v_i$), and set the optimal weight $w_i$ for truthful players. If the platform commits to *any* nonzero weights for the remaining strategic players, the strategic players will always choose their $\alpha_i$ optimally and achieve first-best variance.

## A.4 Extended Numerical Results for Section 3.4

In this section, we extend the analysis of Section 3.4, to a platform with $n = 20$ players, described by a fixed vector $\boldsymbol{\sigma}$ (split evenly across the interval $[0, 1]$, meaning the average signal precision is 0.5). As before, we draw realizations of vector $\boldsymbol{\tau}$ from a uniform in $[0, 1]$ (normalized so that the components sum to 1). For each one of those draws, we record the corresponding value of $\mathcal{G}_\tau$ and $\text{disp}(\boldsymbol{\tau})$.




(a) Optimality gap vs. $\text{disp}(\boldsymbol{\tau})$ (b) $\mathbb{E}$[Optimality gap] vs. $\text{disp}(\boldsymbol{\tau})$ (c) Optimality gap vs. $d(\boldsymbol{\tau}, \boldsymbol{\sigma})$

Figure 6: Optimality gap with heterogeneous tokens and precisions, for $n = 20$.

Figure 6(a) plots the optimality gaps for each mechanism. The $1/n$-weighted optimality gap is unaffected by token dispersion and is represented by the horizontal line. Points located below this line indicate dominance of $\tau$-weighting, and vice versa.

In Figure 6(b), we examine how $\mathcal{G}_\tau$ is affected by token dispersion. As before, we partition $\mathrm{disp}(\boldsymbol{\tau})$ into 50 "bins" of equal width and take the average value of $\mathcal{G}_\tau$ within each one. We also compute the corresponding error bars representing the 95% confidence interval. The results are in line with those of the simpler $n = 2$ player case analyzed in Section 3.4.

Lastly, Figure 6(c) derives additional insights by taking account of the distribution of precisions. From the expression of $V_\tau^{\mathbb{T}}$ in Table 1, $\mathcal{G}_\tau$ depends on the dot product of $\boldsymbol{\tau}$ with $\boldsymbol{\sigma}^2$. Intuitively, the more "misaligned" (orthogonal) vector $\boldsymbol{\tau}$ is to $\boldsymbol{\sigma}^2$, the lower the dot product, and thus, the better token weighting should perform. To capture misalignment between these two vectors, we use the Euclidean distance between them, $d(\boldsymbol{\tau}, \boldsymbol{\sigma})$. Figure 6(c) shows there is a strong negative correlation (-82%), confirming the intuition that the optimality gap under $\tau$-weighting decreases in $d(\boldsymbol{\tau}, \boldsymbol{\sigma})$. The figure also shows that $\tau$-weighting dominates $1/n$-weighting for large enough $d(\boldsymbol{\tau}, \boldsymbol{\sigma})$.

## A.5 Applications of Theorem 1

### A.5.1 Linear information improvement

One particularly meaningful and simple application of Theorem 1 is when players have a linear information improvement function. For instance, let $g(u_i) = 1 + \eta u_i$, where $\eta \geq 0$ captures the effectiveness of how player $i$'s effort translates to improved signal precision. Note, the additive constant equal to 1 is needed so that $g(0) = 1$ is satisfied, that is, when player $i$ exerts no effort, his signal precision is the same as in the base case model of Section 3. In this case, the optimal effort levels are simply

$$u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}) = (f_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}) - 1) \cdot \eta^{-1}. \tag{18}$$

For linear $g$, condition (13) (with $k = 1$) implies $\frac{\partial c(u_i)}{\partial u_i} = \eta(1 + \eta u_i)^2$, and hence, $c(\cdot)$ is a convex positive increasing function. One can readily verify that this type of model satisfies all required conditions. Furthermore, we have already discussed the properties of this equilibrium, in terms of how $u_i^*$ varies with $\tau_i$ and $\rho_i$ (see Proposition 5).

### A.5.2 Concave information improvement

Another case of interest is when the effort improvement function is concave. As discussed, concavity helps to guarantee the existence of an equilibrium, and the comparative statics introduced previously. To this end, consider functions $g(u_i) = (1 + u_i)^\gamma$, where $\frac{1}{3} < \gamma < 1$, then $g$ is concave and $g'(u_i) = \gamma(1 + u_i)^{\gamma - 1}$. Setting $c(u_i) = (1 + u_i)^{3\gamma} - 1$, we have that $c'(u_i)/g'(u_i) = 3g(u_i)^2$, so equation (13) is satisfied, and since $\gamma > \frac{1}{3}$, $c(u_i)$ is convex. The optimal effort levels are given by

$$u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}) = (f_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}))^{1/\gamma} - 1. \tag{19}$$

The implications of this model of concave information improvement are similar to those of the linear model in Section A.5.1; we therefore abstain from repeating them here.

# B Proofs

## B.1 Proofs for Section 3

**Proof of Lemma 1.** Part (i): Our goal is to show that if $\pi$ is well-behaved, (*i.e.,* $\pi(x)$ is symmetric about the origin, and decreasing away from 0), then the function $\mathbb{E}\left[\pi(N(0, \sigma^2)\right]$ is a decreasing function of $\sigma$. To show this, it is sufficient to show that if $0 < \sigma_1 < \sigma_2$, and $X_i \sim N(0, \sigma_i^2)$, are normally distributed random variables, then $\mathbb{E}\left[\pi(X_1)\right] \geq \mathbb{E}\left[\pi(X_2)\right]$.

Let $f_i(x)$ denote the density function of the random variable $X_i$. Since $X_i$ are normally distributed, $f_i(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{x^2}{2\sigma_i^2}}$. To see that $\mathbb{E}\left[\pi(X_1)\right] \geq \mathbb{E}\left[\pi(X_2)\right]$, first, note that

$$\begin{aligned} f_1(x) > f_2(x) \qquad & \text{for } |x| < t \\ f_2(x) > f_1(x) \qquad & \text{for } |x| > t, \end{aligned}$$

where $t = \sqrt{\frac{\log\left(\frac{\sigma_2}{\sigma_1}\right)}{\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}}}$. Note also since both functions have equal measure,

$$\int_0^t (f_1(x) - f_2(x))dx = \int_t^\infty (f_2(x) - f_1(x))dx.$$

We have

$$
\begin{aligned}
\mathbb{E}\left[\pi(X_1)\right] - \mathbb{E}\left[\pi(X_2)\right] &= \int_{-\infty}^{\infty} \pi(x) f_1(x) dx - \int_{-\infty}^{\infty} \pi(x) f_2(x) dx \\
&= \int_{-\infty}^{\infty} \pi(x) \left(f_1(x) - f_2(x)\right) dx \\
&= 2\int_{0}^{\infty} \pi(x) \left(f_1(x) - f_2(x)\right) dx \qquad (\pi(x), f_1(x), f_2(x) \text{ are symmetric}) \\
&= 2\left(\int_{0}^{t} \pi(x) \left(f_1(x) - f_2(x)\right) dx - \int_{t}^{\infty} \pi(t) \left(f_2(x) - f_1(x)\right) dx\right) \\
&\geq 2\left(\int_{0}^{t} \pi(t) \left(f_1(x) - f_2(x)\right) dx - \int_{t}^{\infty} \pi(t) \left(f_2(x) - f_1(x)\right) dx\right) \qquad (\pi(x) \text{ is decreasing}) \\
&= 2\pi(t)\left(\int_{0}^{t} \left(f_1(x) - f_2(x)\right) dx - \int_{t}^{\infty} \left(f_2(x) - f_1(x)\right) dx\right) \\
&= 0.
\end{aligned}
$$

It follows that $\mathbb{E}[\pi(X_1)] \geq \mathbb{E}[\pi(X_2)]$. The inequality becomes strict if $\pi(X)$ is strictly decreasing.
Part (ii): If $f_1(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x}{2\sigma^2}}$ and $f_2(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}$ for some $x_0 > 0$, then

$$
\begin{aligned}
f_1(x) &> f_2(x) \qquad \text{for } x < t \\
f_2(x) &> f_1(x) \qquad \text{for } x > t,
\end{aligned}
$$

for $t = \frac{x_0}{2}$. Now, note that by symmetry $f_1(t-x) - f_2(t-x) = f_2(t+x) - f_1(t+x)$for all $x > 0$. Now, note that since $t > 0$, and $\pi(x)$ is decreasing away from the origin, $\pi(t-x) \geq \pi(t+x)$ for all $x$. Thus

$$
\begin{aligned}
\int_{-\infty}^{\infty} \pi(x) \left(f_1(x) - f_2(x)\right) dx &= \int_{0}^{\infty} \pi(t-x) \left[f_1(t-x) - f_2(t-x)\right] - \pi(t+x) \left[f_2(t+x) - f_1(t+x)\right] dx \\
&\geq 0.
\end{aligned}
$$

A similar result holds when $x_0 < 0$. This shows that it $\mathbb{E}\left[\pi(N(0,\sigma^2))\right] \geq \mathbb{E}\left[\pi(N(x_0,\sigma^2))\right]$ for all $x_0$.
Part (iii): Suppose $\boldsymbol{x}^* \in S$, satisfies $V(\boldsymbol{x}^*) = \min_{x\in S} V(\boldsymbol{x})$. Then for all $\boldsymbol{x} \in S$, $V(\boldsymbol{x}) \geq V(\boldsymbol{x}^*)$, and so by Lemma 1, part (i), $\mathbb{E}\left[\pi\left(N(0, V(\boldsymbol{x}))\right)\right] \leq \mathbb{E}\left[\pi\left(N(0, V(\boldsymbol{x}^*))\right)\right]$. Thus $\mathbb{E}\left[\pi\left(N(0, V(\boldsymbol{x}^*))\right)\right] = \max_{x\in S} \mathbb{E}\left[\pi\left(N(0, V(\boldsymbol{x}))\right)\right]$. In other words, in order to find $\boldsymbol{x}$ that maximizes $\mathbb{E}\left[\pi\left(N(0, V(\boldsymbol{x}))\right)\right]$ subject to the constraint that $\boldsymbol{x} \in S$ it is sufficient to find $\boldsymbol{x} \in S$ that minimizes $V(\boldsymbol{x})$. □

**Proof of Proposition 1.** First, note that since the aggregation function is linear, $\hat{q}$ is a linear combination of the $v_j$, and therefore $q - \hat{q}$ is normally distributed with

$$\mathbb{E}\left[q - \hat{q}\right] = \mu - \sum_{j=1}^{n} \tau_j \left(\alpha_j \, \mathbb{E}(s_j) + (1-\alpha_j)\mu\right) = 0,$$

and

$$\begin{aligned} \text{Var}\left[q - \hat{q}\right] &= \text{Var}\left[q - \sum_{j=1}^{n} \tau_j \alpha_j s_j\right] \\ &= \text{Var}\left[\left(1 - \sum_{j=1}^{n} \tau_j \alpha_j\right) q - \sum_{j=1}^{n} \tau_j \alpha_j \epsilon_j\right] \\ &= \left(1 - \sum_{j=1}^{n} \tau_j \alpha_j\right)^2 \sigma_q^2 + \sum_{j=1}^{n} \tau_j^2 \alpha_j^2 \sigma_j^2. \end{aligned} \tag{20}$$

Now, consider that given the restriction to linear strategies $v_i = \alpha_i s_i + (1-\alpha_i)\mu$, player $i$'s strategy is completely determined by $\alpha_i$. It is well-known that in linear models with Gaussian signals, when all players are playing linear strategies, player $i$'s best response is linear as well (Myatt and Wallace 2012), implying that the choice of $\alpha_i$ does not depend on the realization of the signal $s_i$. It follows that player $i$'s optimization problem in (4) can be written, unconditionally, as

$$\max_{\alpha_i} \mathbb{E}\left[\pi_i \left(q - \hat{q}\right)\right], \tag{21}$$

where player $i$'s payoff is $\pi_i \left(q - \hat{q}\right) = \tau_i \cdot \pi \left(q - \hat{q}\right)$. Assuming $\tau_i > 0$, maximizing $\mathbb{E}[\pi_i \left(q - \hat{q}\right)]$ is equivalent to maximizing $\mathbb{E}[\pi \left(q - \hat{q}\right)]$. Since $\pi$ is well behaved and $q - \hat{q}$ is normally distributed with mean 0, Lemma 1 applies, and therefore choosing $\alpha_i$ to maximize $\mathbb{E}\left[\pi_i(q - \hat{q})\right]$ is equivalent to choosing $\alpha_i$ to minimize $\text{Var}\left[q - \hat{q}\right]$. Now, $\text{Var}\left[q - \hat{q}\right]$ is a convex quadratic function of $\alpha_i$, and thus has a unique minimum. Taking derivatives with respect to $\alpha_i$, we have

$$\frac{\partial}{\partial \alpha_i} \left(1 - \sum_{j=1}^{n} \tau_j \alpha_j\right)^2 \sigma_q^2 + \sum_{j=1}^{n} \tau_j^2 \alpha_j^2 \sigma_j^2 = -2 \left(1 - \sum_{j=1}^{n} \tau_j \alpha_j\right) \tau_i \sigma_q^2 + 2\tau_i^2 \alpha_i \sigma_i^2.$$

The first-order condition is thus

$$-2 \left(1 - \sum_{j=1}^{n} \tau_j \alpha_j\right) \tau_i \sigma_q^2 + 2\tau_i^2 \alpha_i \sigma_i^2 = 0,$$

which becomes

$$\left(1-\sum_{j=1}^{n}\tau_j\alpha_j\right)\sigma_q^2=\alpha_i\tau_i\sigma_i^2. \tag{22}$$

Notice that the left-hand side of equation (22) is fixed for all $i$, therefore, in equilibrium, it must be that

$$\alpha_i\tau_i\sigma_i^2=\alpha_1\tau_1\sigma_1^2,$$

for all $i$. Thus $\alpha_i=\frac{\alpha_1\tau_1\sigma_1^2}{\tau_i\sigma_i^2}$. Plugging this back into equation (22), we have

$$\left(1-\alpha_1\tau_1\sigma_1^2\sum_{j=1}^{n}\sigma_j^{-2}\right)\sigma_q^2=\alpha_1\tau_1\sigma_1^2. \tag{23}$$

Then, solving equation (23) for $\alpha_1$, we have

$$\sigma_q^2=\alpha_1\tau_1\sigma_1^2\left(1+\sigma_q^2\sum_{j=1}^{n}\sigma_j^{-2}\right),$$

which implies that in equilibrium

$$\alpha_j^*=\frac{\sigma_q^2}{\tau_j\sigma_j^2\left(1+\sigma_q^2\sum_{j=1}^{n}\sigma_j^{-2}\right)}=\frac{1}{\tau_j\sigma_j^2\left(\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}\right)}. \tag{24}$$

Recall that from (20),

$$\mathrm{Var}\left[q-\hat{q}\right]=\left(1-\sum_{j=1}^{n}\tau_j\alpha_j\right)^2\sigma_q^2+\sum_{j=1}^{n}\tau_j^2\alpha_j^2\sigma_j^2.$$

Now, notice that

$$\tau_j\alpha_j^*=\frac{1}{\sigma_j^2}\cdot\left(\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}\right)^{-1},$$

which implies

$$1-\sum_{j=1}^{n}\tau_j\alpha_j^*=\frac{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}}{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}}-\sum_{j=1}^{n}\frac{\sigma_j^{-2}}{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}}=\frac{\sigma_q^{-2}}{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}}.$$

Solving for the variance in equilibrium, we have

$$\begin{aligned}
\operatorname{Var}\left[q-\hat{q}\right] &= \left(1-\sum_{j=1}^{n}\tau_j\alpha_j^*\right)^2 \sigma_q^2 + \sum_{j=1}^{n}\tau_j^2(\alpha_j^*)^2\sigma_j^2 \\
&= \left(\frac{\sigma_q^{-2}}{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}}\right)^2 \sigma_q^2 + \sum_{j=1}^{n}\tau_j^2(\alpha_j^*)^2\sigma_j^2 \\
&= \left(\frac{\sigma_q^{-2}}{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}}\right)^2 \sigma_q^2 + \sum_{j=1}^{n}\frac{\sigma_j^{-2}}{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}} \\
&= \frac{1}{\sigma_q^{-2}+\sum_{j=1}^{n}\sigma_j^{-2}}. \qquad (25)
\end{aligned}$$

Note, these results are generalizable to symmetric distributions beyond normal, but require a different proof methodology. □

**Proof of Proposition 2.** Recall $\mathbb{T}, \mathbb{S}$ represent the sets of truthful and strategic players, respectively, such that $|\mathbb{T} \cup \mathbb{S}| = n$. Player votes are given by

$$v_i = \begin{cases} \frac{\sigma_q^2}{\sigma_q^2+\sigma_i^2} s_i + \left(1 - \frac{\sigma_q^2}{\sigma_q^2+\sigma_i^2}\right)\mu & \text{for } i \in \mathbb{T} \\ \alpha_i s_i + (1-\alpha_i)\mu & \text{for } i \in \mathbb{S}\ . \end{cases}$$

By definition, the strategies of truthful players are fixed, i.e., independent of the actions of all other players. In this setting, with mixed players,

$$\begin{aligned}
\mathbb{E}\left[q-\hat{q}\right] &= \mu - \sum_{j\in\mathbb{T}}\tau_j\left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\,\mathbb{E}(s_j) + (1-\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2})\mu\right) - \sum_{j\in\mathbb{S}}\tau_j\left(\alpha_j\,\mathbb{E}(s_j)+(1-\alpha_j)\mu\right) \\
&= \mu - \sum_{j\in\mathbb{T}}\tau_j\mu - \sum_{j\in\mathbb{S}}\tau_j\mu \\
&= 0,
\end{aligned}$$

and $\mathrm{Var}\left[q-\hat{q}\right]$ is given by

$$\begin{aligned}
&\mathrm{Var}\left[q-\sum_{j\in\mathbb{T}}\tau_j\left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}s_j+(1-\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2})\mu\right)-\sum_{j\in\mathbb{S}}\tau_j\left(\alpha_j s_j+(1-\alpha_j)\mu\right)\right]\\
&=\mathrm{Var}\left[q-\sum_{j\in\mathbb{T}}\tau_j\left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}s_j\right)-\sum_{j\in\mathbb{S}}\tau_j\left(\alpha_j s_j\right)\right]\\
&=\mathrm{Var}\left[\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}-\sum_{j\in\mathbb{S}}\tau_j\alpha_j\right)q-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\epsilon_j-\sum_{j\in\mathbb{S}}\tau_j\alpha_j\epsilon_j\right]\\
&=\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}-\sum_{j\in\mathbb{S}}\tau_j\alpha_j\right)^2\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2+\sum_{j\in\mathbb{S}}\left(\tau_j\alpha_j\right)^2\sigma_j^2.
\end{aligned}$$

Player $i$'s payoff is $\pi_i(q-\hat{q})=\tau_i\cdot\pi(q-\hat{q})$ and from Lemma 1, choosing $\alpha_i$ to maximize $\mathbb{E}\left[\pi_i(q-\hat{q})\right]$ is equivalent to choosing $\alpha_i$ to minimize $\mathrm{Var}\left[q-\hat{q}\right]$. Now, $\mathrm{Var}[q-\hat{q}]$ is a convex quadratic function of $\alpha_i$, and thus has a unique minimum. Taking derivatives with respect to $\alpha_i$, we have

$$\begin{aligned}
&\frac{\partial}{\partial\alpha_i}\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}+\sum_{j\in\mathbb{S}}\tau_j\alpha_j\right)^2\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2+\sum_{j\in\mathbb{S}}\left(\tau_j\alpha_j\right)^2\sigma_j^2\\
&=-2\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}-\sum_{j\in\mathbb{S}}\tau_j\alpha_j\right)\tau_i\sigma_q^2+2\tau_i^2\sigma_i^2\alpha_i.
\end{aligned}$$

The first-order condition is thus

$$-2\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}-\sum_{j\in\mathbb{S}}\tau_j\alpha_j\right)\tau_i\sigma_q^2+2\tau_i^2\sigma_i^2\alpha_i=0,$$

which becomes

$$\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}-\sum_{j\in\mathbb{S}}\tau_j\alpha_j\right)\sigma_q^2=\tau_i\alpha_i\sigma_i^2 \tag{26}$$

Since the left-hand side of equation (26) is the same for all $i$, in equilibrium, it must be that

$$\alpha_i\tau_i\sigma_i^2=\alpha_1\tau_1\sigma_1^2,$$

for all $i$. Thus $\alpha_i = \frac{\alpha_1 \tau_1 \sigma_1^2}{\tau_i \sigma_i^2}$. Plugging this back into equation (26), we have

$$\left(1 - \sum_{j\in\mathbb{T}} \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} - \sum_{j\in\mathbb{S}} \tau_j \frac{\alpha_1 \tau_1 \sigma_1^2}{\tau_j \sigma_j^2}\right) \sigma_q^2 = \tau_1 \alpha_1 \sigma_1^2 \tag{27}$$

Solving equation (27) for $\alpha_1$, we have

$$\left(1 - \sum_{j\in\mathbb{T}} \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2}\right) \sigma_q^2 = \alpha_1 \left(\tau_1 \sigma_1^2 + \sum_{j\in\mathbb{S}} \tau_j \frac{\tau_1 \sigma_1^2}{\tau_j \sigma_j^2} \sigma_q^2\right) = \alpha_1 \sigma_1^2 \tau_1 \left(1 + \sum_{j\in\mathbb{S}} \frac{\sigma_q^2}{\sigma_j^2}\right)$$

which implies that in equilibrium

$$\alpha_i^* = \frac{\left(1 - \sum_{j\in\mathbb{T}} \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2}\right) \sigma_q^2}{\sigma_i^2 \tau_i \left(1 + \sum_{j\in\mathbb{S}} \frac{\sigma_q^2}{\sigma_j^2}\right)}. \tag{28}$$

Next, we derive the corresponding platform variance. Define

$$A \stackrel{\text{def}}{=} 1 - \sum_{j\in\mathbb{T}} \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} \tag{29}$$

$$B \stackrel{\text{def}}{=} 1 + \sum_{j\in\mathbb{S}} \frac{\sigma_q^2}{\sigma_j^2}. \tag{30}$$

Thus

$$\tau_i \alpha_i^* = \frac{A \sigma_q^2}{B \sigma_i^2}$$

$$\sum_{j\in\mathbb{S}} \tau_j \alpha_j^* = \sum_{j\in\mathbb{S}} \frac{A \sigma_q^2}{B \sigma_j^2} = \frac{A}{B} (B - 1)$$

$$\sum_{j\in\mathbb{S}} (\tau_j \alpha_j^*)^2 \sigma_j^2 = \sum_{j\in\mathbb{S}} \frac{A^2 \sigma_q^4}{B^2 \sigma_j^2} = \frac{A^2 \sigma_q^2}{B^2} (B - 1).$$

Starting from $\mathrm{Var}[q - \hat{q}]$ as

$$\left(1 - \sum_{j\in\mathbb{T}} \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} - \sum_{j\in\mathbb{S}} \tau_j \alpha_j\right)^2 \sigma_q^2 + \sum_{j\in\mathbb{T}} \left(\tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2}\right)^2 \sigma_j^2 + \sum_{j\in\mathbb{S}} (\tau_j \alpha_j)^2 \sigma_j^2,$$

we can rewrite this expression as follows:

$$
\begin{aligned}
&\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}-\sum_{j\in\mathbb{S}}\tau_j\alpha_j^*\right)^2\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2+\sum_{j\in\mathbb{S}}\left(\tau_j\alpha_j^*\right)^2\sigma_j^2\\
&=\left(A-\sum_{j\in\mathbb{S}}\tau_j\alpha_j^*\right)^2\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2+\sum_{j\in\mathbb{S}}\left(\tau_j\alpha_j^*\right)^2\sigma_j^2\\
&=\left(A-\left(\frac{A}{B}(B-1)\right)\right)^2\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2+\sum_{j\in\mathbb{S}}\left(\tau_j\alpha_j^*\right)^2\sigma_j^2\\
&=\left(A-\left(\frac{A}{B}(B-1)\right)\right)^2\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2+\frac{A^2\sigma_q^2}{B^2}(B-1)\\
&=\left(\frac{A}{B}\right)^2\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2+\frac{A^2\sigma_q^2}{B^2}(B-1)\\
&=\frac{A^2}{B}\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2,
\end{aligned}
$$

which, after replacing the expressions for $A$ and $B$, gives

$$
\begin{aligned}
V^{\mathbb{T},\mathbb{S}}&=\frac{\left(1-\sum_{j\in\mathbb{T}}\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2}{1+\sum_{j\in\mathbb{S}}\frac{\sigma_q^2}{\sigma_j^2}}\sigma_q^2+\sum_{j\in\mathbb{T}}\left(\tau_j\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2\sigma_j^2\\
&=\left(1-\sum_{j\in\mathbb{T}}\tau_j\beta_j\right)^2V^{fb}+\sum_{j\in\mathbb{T}}\left(\tau_j\beta_j\sigma_j\right)^2\\
&\geq V^{fb},\quad\text{if}\quad|\mathbb{T}|\geq 1.
\end{aligned}
$$

□

**Proof of Proposition 3.** Consider a platform with $n$ players, such that $|\mathbb{T}|>0$. To show that the platform's payoff increases in the number of strategic players, it suffices to show that if any truthful player is randomly chosen and turned into a strategic player, the platform's payoff increases.

By Lemma 1, to show that the equilibrium payoff increases, it is sufficient to show that the variance in equilibrium decreases. Suppose player $j$ moves from truthful to strategic. Let $V_0$ denote the equilibrium variance when player $j$ was truthful, and $V_1$ denote the equilibrium variance once player $j$ becomes strategic. Similarly, let $A_0, B_0$ be the quantities (defined in Equations 29, 30) when player $j$ is truthful, and $A_1, B_1$ be the same equations once player $j$ is strategic. With this notation, our goal is to show that $V_0 - V_1 \geq 0$.

Recall

$$V = \frac{A^2}{B}\sigma_q^2 + \sum_{j \in \mathbb{T}} \left( \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} \right)^2 \sigma_j^2. \tag{31}$$

Now, note that $A_1 = A_0 - p_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}$ $B_1 = B_0 + \frac{\sigma_q^2}{\sigma_j^2}$.

Then

$$\begin{aligned}
V_0 - V_1 &= \sigma_q^2 \left( \frac{A_0^2}{B_0} - \frac{A_1^2}{B_1} \right) + \left( \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} \right)^2 \sigma_j^2 \\
&= \sigma_q^2 \left( \frac{B_1 A_0^2 - B_0 A_1^2}{B_0 B_1} \right) + \left( \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} \right)^2 \sigma_j^2 \\
&= \sigma_q^2 \left( \frac{\left( B_0 + \frac{\sigma_q^2}{\sigma_j^2} \right) A_0^2 - B_0 \left( A_0 - \tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2} \right)^2}{B_0 \left( B_0 + \frac{\sigma_q^2}{\sigma_j^2} \right)} \right) + \left( \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} \right)^2 \sigma_j^2 \\
&= \sigma_q^2 \left( \frac{B_0 A_0^2 + 2 B_0 A_0 \tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2} + A_0^2 \frac{\sigma_q^2}{\sigma_j^2} - B_0 A_0^2 - B_0 \tau_j^2 \left( \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2} \right)^2}{B_0 \left( B_0 + \frac{\sigma_q^2}{\sigma_j^2} \right)} \right) + \left( \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} \right)^2 \sigma_j^2 \\
&= \sigma_q^2 \left( \frac{2 B_0 A_0 \tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2} + A_0^2 \frac{\sigma_q^2}{\sigma_j^2} - B_0 \tau_j^2 \left( \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2} \right)^2}{B_0 \left( B_0 + \frac{\sigma_q^2}{\sigma_j^2} \right)} \right) + \left( \tau_j \frac{\sigma_q^2}{\sigma_q^2 + \sigma_j^2} \right)^2 \sigma_j^2.
\end{aligned}$$

Next, notice the last line is

$$\geq \sigma_q^2 \left( \frac{2B_0 A_0 \tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2} - B_0 \tau_j^2 \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2}{B_0 \left(B_0 + \frac{\sigma_q^2}{\sigma_j^2}\right)} \right) + \left(\tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2 \sigma_j^2$$

$$= \sigma_q^2 \left( \frac{2A_0 \tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2} - \tau_j^2 \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2}{B_0 + \frac{\sigma_q^2}{\sigma_j^2}} \right) + \left(\tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2 \sigma_j^2$$

$$\geq \sigma_q^2 \left( \frac{-\tau_j^2 \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2}{B_0 + \frac{\sigma_q^2}{\sigma_j^2}} \right) + \left(\tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2 \sigma_j^2$$

$$= \frac{-\tau_j^2 \sigma_j^2 \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2}{B_0 \frac{\sigma_j^2}{\sigma_q^2} + 1} + \left(\tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2 \sigma_j^2$$

$$\geq -\tau_j^2 \sigma_j^2 \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2 + \left(\tau_j \frac{\sigma_q^2}{\sigma_q^2+\sigma_j^2}\right)^2 \sigma_j^2 = 0,$$

which completes the proof. □

**Proof of Proposition 4.** Recall, the optimality gap is defined as

$$\mathcal{G}_j = V_j^{\mathbb{T}} - V^{\mathbb{S}}, \quad j \in \{\tau, 1/n\}.$$

It follows that $\mathcal{G}_\tau^{\mathbb{T}} - \mathcal{G}_{1/n}^{\mathbb{T}} = V_\tau^{\mathbb{T}} - V_{1/n}^{\mathbb{T}}$. It thus suffices to show the results for $V_\tau^{\mathbb{T}} - V_{1/n}^{\mathbb{T}}$.

Part (i): In the symmetric information case, the expressions from Table 1 simplify to $V_\tau^{\mathbb{T}} = (\beta\sigma)^2 \sum_{i=1}^n \tau_i^2 + \sigma_q^2(1-\beta)^2$ and $V_{1/n}^{\mathbb{T}} = (\beta\sigma)^2 \sum_{i=1}^n \frac{1}{n^2} + \sigma_q^2(1-\beta)^2$. Taking the difference, we obtain

$$\left(V^{\mathbb{T}} - V_{1/n}^{\mathbb{T}}\right) \propto \sum_{i=1}^n \left(\tau_i^2 - \frac{1}{n^2}\right), \tag{32}$$

and hence the difference has the same sign as this latter term. Considering $\sum_{i=1}^n \tau_i = 1$, we must have $\sum_{i=1}^n \left(\tau_i^2 - \frac{1}{n^2}\right) \geq 0$ irrespective of how the $\tau_i$'s are distributed. This is because $\min_{\tau_1,\ldots,\tau_n} \sum_{i=1}^n \tau_i^2$ subject to $\sum_{i=1}^n \tau_i = 1$, is convex and attains its minimum at $\tau_i^* = \frac{1}{n}, \forall i$. To see this, e.g., take the FOC of the Lagrangian $\sum_{i=1}^n \tau_i^2 + \lambda \left(\sum_{i=1}^n \tau_i - 1\right)$. The difference being $\geq 0$ implies $V_\tau^{\mathbb{T}} \geq V_{1/n}^{\mathbb{T}}$.

Part (ii): We have

$$\tau_i^2 - \frac{1}{n^2} = \tau_i^2 - 2\frac{1}{n^2} + \frac{1}{n^2} - 2\frac{\tau_i}{n} + 2\frac{\tau_i}{n} = \left(\tau_i - \frac{1}{n}\right)^2 - 2\frac{1}{n^2} + 2\frac{\tau_i}{n}.$$

Hence,

$$\begin{aligned}\sum_{i=1}^n \left(\tau_i^2 - \frac{1}{n^2}\right) = \sum_{i=1}^n \left\{\left(\tau_i - \frac{1}{n}\right)^2 - 2\frac{1}{n^2} + 2\frac{\tau_i}{n}\right\} &= \sum_{i=1}^n \left(\tau_i - \frac{1}{n}\right)^2 - 2\frac{1}{n} + 2\frac{1}{n}\sum_{i=1}^n \tau_i \\ &= \sum_{i=1}^n \left(\tau_i - \frac{1}{n}\right)^2 \\ &= \sum_{i=1}^n (\tau_i - \bar{\tau})^2 \\ &= ||\boldsymbol{\tau} - \bar{\boldsymbol{\tau}}||^2\end{aligned}$$

where the second-to-last line follows from $\bar{\tau} = \frac{1}{n}\sum_{i=1}^n \tau_i = \frac{1}{n}$. Given (32) and $n > 0$, $\left(V_\tau^{\mathbb{T}} - V_{1/n}^{\mathbb{T}}\right)$ is increasing in $||\boldsymbol{\tau} - \bar{\boldsymbol{\tau}}|| \stackrel{\text{def}}{=} \text{disp}(\boldsymbol{\tau})$. □

## B.2 Proofs for Section 4

**Proof of Lemma 2.** Using the same arguments as those provided in the proof of Proposition 1, the restriction to linear voting strategies implies the second-stage maximization problem is equivalent to the following optimization problem over signal weights $\alpha_i$:

$$\max_{\alpha_i} \left\{\mathbb{E}\left[k_1 - k_2\big(q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{\alpha}))\big)^2\right]\right\} = k_1 - k_2 \min_{\alpha_i} \text{Var}\left[q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{\alpha}))\right],$$

where the equality follows from Lemma 1 and the fact that $\mathbb{E}[(q - \hat{q})^2] = \text{Var}[q - \hat{q}]$ given that the signals are unbiased (and thus $\mathbb{E}[q - \hat{q}] = 0$).

Because $\boldsymbol{u}$ is fixed in this second stage, the objective $\min_{\alpha_i} \text{Var}[q - \hat{q}]$ is equivalent to the one solved in Proposition 1, after a simple change of variables. Specifically, let $\rho' = \rho g(u_i) = \frac{g(u_i)}{\sigma_i^2} = \frac{1}{\sigma_i'^2}$. By identification using (24) and (25) in the proof of Proposition 1, we have

$$\alpha_i^*(\boldsymbol{u}) \stackrel{\text{def}}{=} \text{argmin}_{\alpha_i} \text{Var}\left[q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{\alpha}))\right] = \frac{1}{(\rho_i')^{-1}\tau_i\left(\rho_q + \sum_{j=1}^n \rho_j'\right)} = \frac{\rho_i g(u_i)}{\tau_i\left(\rho_q + \sum_{j=1}^n \rho_j g(u_j)\right)}.$$

Letting $\alpha^*(\boldsymbol{u})$ be the vector with components $\alpha_i^*(\boldsymbol{u}), i \in \{1, \ldots, n\}$, we also have

$$V(\boldsymbol{u}) \stackrel{\text{def}}{=} \operatorname{Var}\left[q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{\alpha}^*(\boldsymbol{u})))\right] = \left(\rho_q + \sum_{j=1}^{n} \rho_j'\right)^{-1} = \left(\rho_q + \sum_{j=1}^{n} \rho_j g(u_j)\right)^{-1}.$$

As a side remark, given the above results follow from those of Proposition 1, we also have that players would be able to achieve $V(\boldsymbol{u})$ even under alternative aggregation mechanisms, e.g., if the platform had chosen $1/n$-weighted aggregation. To see this, replace $\tau_i \to 1/n$ in the expression for $\alpha_i^*(\boldsymbol{u})$, and recompute $V(\boldsymbol{u})$ to obtain the same form.

Using these results, for a fixed $\boldsymbol{u}$, the second stage problem is equivalent to

$$\max_{\alpha_i} \left\{\mathbb{E}\left[k_1 - k_2\big(q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{\alpha}))\big)^2\right]\right\} = k_1 - k_2 \operatorname{Var}\left[q - \hat{q}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{\alpha}^*(\boldsymbol{u})))\right] = k_1 - k_2 V(\boldsymbol{u}).$$

In turn, the first stage objective simplifies to

$$\max_{u_i \geq 0} -c(u_i) + \tau_i(k_1 - k_2 V(\boldsymbol{u})) = \tau_i k_1 + \max_{u_i \geq 0} -c(u_i) - \tau_i k_2 V(\boldsymbol{u}) = \tau_i k_1 - \min_{u_i \geq 0} c(u_i) + \tau_i k_2 V(\boldsymbol{u}).$$

Setting arbitrary constants $k_1 = 0$ and $k_2 = 1$ gives the desired result.

□

**Proof of Lemma 3.** A direct calculation gives

$$V' = \frac{\partial V(\boldsymbol{u})}{\partial u_i} = -\frac{\rho_i g'(u_i)}{\rho^2(\boldsymbol{u})}, \text{ and we have } V' \leq 0 \text{ given } g'(u_i) > 0, \tag{33}$$

$$V'' = \frac{\partial^2 V(\boldsymbol{u})}{\partial u_i^2} = \frac{\rho_i \left(2\rho_i g'(u_i)^2 - g''(u_i)\rho(\boldsymbol{u})\right)}{\rho(\boldsymbol{u})^3} = \rho_i V^2(\boldsymbol{u}) \left(2\rho_i g'(u_i)^2 V(\boldsymbol{u}) - g''(u_i)\right), \tag{34}$$

where in the last equation, we use the identity $\rho(\boldsymbol{u})^{-1} = V(\boldsymbol{u})$. By construction, $\rho_i, \rho(\boldsymbol{u}), V(\boldsymbol{u}) \geq 0$, hence $V$ is convex iff $2\rho_i g'(u_i)^2 - g''(u_i)\rho(\boldsymbol{u}) \geq 0$, which simplifies to $\frac{\rho_i}{\rho(\boldsymbol{u})} \geq \frac{1}{2}\frac{g''(u_i)}{g'(u_i)^2}$, or alternatively, $\rho_i V(\boldsymbol{u}) \geq \frac{1}{2}\frac{g''(u_i)}{g'(u_i)^2}$. □

**Proof of Theorem 1.** Suppose the payoff is quadratic, *i.e.,* the payoff is $k_1 - (q - \hat{q})^2$. Then the expected payoff for player $i$ is

$$\tau_i(k_1 - V(\boldsymbol{u})) - c(u_i).$$

Thus,

$$\frac{\partial}{\partial u_i}\left(\tau_i(k_1 - V(\boldsymbol{u})) - c(u_i)\right) = -\tau_i \frac{\partial}{\partial u_i} V(\boldsymbol{u}) - c'(u_i)$$
$$= \tau_i V(\boldsymbol{u})^2 \cdot g'(u_i)\rho_i - c'(u_i),$$

and

$$\frac{\partial^2}{\partial u_i^2}\left(\tau_i(k_1 - V(\boldsymbol{u})) - c(u_i)\right) = \frac{\partial}{\partial u_i}\left(\tau_i \rho_i V(\boldsymbol{u})^2 \cdot g'(u_i) - c'(u_i)\right)$$
$$= -2\tau_i \rho_i^2 V(\boldsymbol{u})^3 \left(g'(u_i)\right)^2 + \tau_i \rho_i V(\boldsymbol{u})^2 g''(u_i) - c''(u_i).$$

In equilibrium, the first derivative of the payoff is zero, which means

$$\tau_i \rho_i V(\boldsymbol{u})^2 \cdot g'(u_i) = c'(u_i) \Rightarrow \tau_i \rho_i V(\boldsymbol{u})^2 = \frac{c'(u_i)}{g'(u_i)}. \tag{35}$$

Since this holds for all $i$, if we multiply two such equations, and cancel out the $V$s, we have

$$\tau_i \rho_i \frac{c'(u_j)}{g'(u_j)} = \tau_j \rho_j \frac{c'(u_i)}{g'(u_i)} \Rightarrow \frac{\tau_i \rho_i}{\tau_1 \rho_1}\left(\frac{c'(u_1)}{g'(u_1)}\right) = \left(\frac{c'(u_i)}{g'(u_i)}\right). \tag{36}$$

Suppose $c, g$ satisfy Assumption 3, *i.e.*, $\frac{c'(u_i)}{g'(u_i)} = k \cdot g(u_i)^2$, for some constant $k$. Then, equation (36) becomes

$$g(u_i) = \sqrt{\frac{\tau_i \rho_i}{\tau_1 \rho_1}} g(u_1). \tag{37}$$

Further, equation (35) becomes

$$\tau_1 \rho_1 V(\boldsymbol{u})^2 = k g(u_1)^2,$$

which means

$$\sqrt{\frac{\tau_1 \rho_1}{k}} = g(u_1) V(\boldsymbol{u})^{-1}. \tag{38}$$

Now,

$$V^{-1}(\boldsymbol{u}) \stackrel{\text{def}}{=} \rho(\boldsymbol{u}) = \rho_q + \sum_{j=1}^{n} g(u_j)\rho_j.$$

Plugging in $g(\cdot)$ from (37) we have

$$\rho(\boldsymbol{u}) = \rho_q + g(u_1) \sum_{i=j}^{n} \sqrt{\frac{\tau_j \rho_j^3}{\tau_1 \rho_1}}.$$

Next, replacing this in equation (38) obtains

$$\sqrt{\frac{\tau_1\rho_1}{k}} = g(u_1)\left(\rho_q + g(u_1)\sum_{j=1}^{n}\sqrt{\frac{\tau_j\rho_j^3}{\tau_1\rho_1}}\right).$$

Rearranging, we have the quadratic:

$$g(u_1)^2\sum_{j=1}^{n}\sqrt{\frac{\tau_j\rho_j^3}{\tau_1\rho_1}} + g(u_1)\rho_q - \sqrt{\frac{\tau_1\rho_1}{k}}.$$

By the quadratic equation, this has solutions

$$g(u_1) = \frac{-\rho_q \pm \sqrt{\rho_q^2 + \frac{4}{k}\sum_{j=1}^{n}\sqrt{\tau_j\rho_j^3}}}{\frac{2}{\sqrt{\tau_1\rho_1}}\sum_{j=1}^{n}\sqrt{\tau_j\rho_j^3}}.$$

Only the first solution can be positive, thus we have that in equilibrium

$$g(u_1) = \frac{-\rho_q + \sqrt{\rho_q^2 + \frac{4}{k}\sum_{j=1}^{n}\sqrt{\tau_j\rho_j^3}}}{\frac{2}{\sqrt{\tau_1\rho_1}}\sum_{j=1}^{n}\sqrt{\tau_j\rho_j^3}}.$$

Finally, by equation (37), we obtain the solution for arbitrary $i$

$$g(u_i) = \sqrt{\frac{\tau_i\rho_i}{\tau_1\rho_1}}g(u_1) = \frac{-\rho_q + \sqrt{\rho_q^2 + \frac{4}{k}\sum_{j=1}^{n}\sqrt{\tau_j\rho_j^3}}}{\frac{2}{\sqrt{\tau_i\rho_i}}\sum_{j=1}^{n}\sqrt{\tau_j\rho_j^3}}.$$

Assuming $g$ is invertible gives the desired result. □

**Proof of Proposition 5.** We first prove parts (i), (iii), (iv), and then part (ii).

In all cases, $g(\cdot)$ increasing (Assumption 2) $\Rightarrow g^{-1}(\cdot)$ increasing, and thus the sign of differences in effort levels $u_i$, is given by that of differences in $f_i$, defined in (14). It thus suffices to consider the comparative statics of $f_i$. (We suppress superscripts and some dependent variables to ease exposition).

*Part (i):* Here, we consider how a change in $\tau_i$ affects $f_i$. One subtlety is that $\tau_i$ is constrained via the equation $\tau_1 + \cdots + \tau_n = 1$, and thus increasing $\tau_i$ implies (weakly) decreasing $\tau_j, j \neq i$.[14]

[14]To see this, consider the definition of relative token holdings as a function of absolute token holdings, i.e., $\tau_i = \frac{n_i}{\sum_{j=1}^{n} n_j}, i \in \{1, \ldots, n\}$, where $n_i$ represents the absolute number of tokens held by player $i$. Within this context,

Suppose then that $\boldsymbol{\tau} \stackrel{\text{def}}{=} \{\tau_1, \ldots, \tau_n\}$ and $\boldsymbol{\tau'} \stackrel{\text{def}}{=} \{\tau'_1, \ldots, \tau'_n\}$ are two vectors such that $\tau_i \geq \tau'_i$ and $\tau_j \leq \tau'_j, \forall j \neq i$. We want to show this implies $f_i(\boldsymbol{\tau}, \rho) \geq f_i(\boldsymbol{\tau'}, \boldsymbol{\rho})$.

Let

$$x \stackrel{\text{def}}{=} \sum_{j=1}^{n} \left(\tau_j \rho_j^3\right)^{1/2}, \quad x' \stackrel{\text{def}}{=} \sum_{j=1}^{n} \left(\tau'_j \rho_j^3\right)^{1/2}, \tag{39}$$

and note that with this notation, the expression in (14) is written as $f_i(\boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau_i)^{1/2}} x}$.

Because a change from $\tau_i$ to $\tau'_i$ might result in the sums $x \geq x'$ or $x \leq x'$ (depending on the distribution of precisions), we need to consider both cases.

Case 1: $x \geq x'$. In this case

$$f_i(\boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau_i)^{1/2}} \sum_{j=1}^{n} \left(\tau_j \rho_j^3\right)^{1/2}} \geq \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x'\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau_i)^{1/2}} \sum_{j=1}^{n} \left(\tau_j \rho_j^3\right)^{1/2}} = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x'\right)^{1/2} - \rho_q}{2\frac{1}{\rho_i^{1/2}} \sum_{j=1}^{n} \left(\frac{\tau_j}{\tau_i} \rho_j^3\right)^{1/2}} \stackrel{\text{def}}{=} y$$

Here, notice that $\tau_i \geq \tau'_i$ and $\tau_j \leq \tau'_j, \forall j \neq i \Rightarrow \frac{\tau_j}{\tau_i} \leq \frac{\tau'_j}{\tau'_i}, \forall j \in \{1, \ldots, n\}$, and thus, the denominator can be bounded $\sum_{j=1}^{n} \left(\frac{\tau_j}{\tau_i} \rho_j^3\right)^{1/2} \leq \sum_{j=1}^{n} \left(\frac{\tau'_j}{\tau'_i} \rho_j^3\right)^{1/2}$. We thus have

$$y \geq \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x'\right)^{1/2} - \rho_q}{2\frac{1}{\rho_i^{1/2}} \sum_{j=1}^{n} \left(\frac{\tau'_j}{\tau'_i} \rho_j^3\right)^{1/2}} = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x'\right)^{1/2} - \rho_q}{2\frac{1}{(\tau'_i \rho_i)^{1/2}} x'} = f_i(\boldsymbol{\tau'}, \boldsymbol{\rho}).$$

Case 2: $x' \geq x$. In this case

$$f_i(\boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau_i)^{1/2}} x} \geq \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau'_i)^{1/2}} x}.$$

Thus to show that $f_i(\boldsymbol{\tau}, \boldsymbol{\rho}) \geq f_i(\boldsymbol{\tau'}, \boldsymbol{\rho})$ it suffices to show that $\frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau'_i)^{1/2}} x} \geq \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}} x'\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau'_i)^{1/2}} x'} = f_i(\boldsymbol{\tau'}, \boldsymbol{\rho})$, for $x' \geq x$. This holds if the function $F(x) \stackrel{\text{def}}{=} \frac{\left(a^2 + bx\right)^{1/2} - a}{x}$, with $a \stackrel{\text{def}}{=} \rho_q \geq 0, b \stackrel{\text{def}}{=} \frac{4}{k^{1/2}} \geq 0$,

increasing $\tau_i$ can be interpreted as increasing $n_i$, i.e., player $i's$ absolute token holdings. But, an increase in $n_i$ also implies an increase in the denominator $\sum_{j=1}^{n} n_j$ and thus, a strict decrease in all non-zero $\tau_j = \frac{n_j}{\sum_{j=1}^{n} n_j}, j \neq i$.

is decreasing, i.e., if $F'(x) \leq 0, x \geq 0$. We have

$$F'(x) = \frac{1}{2x^2}\left(2a - \frac{2a^2 + bx}{\sqrt{a^2 + bx}}\right).$$

Let $G(b) = 2a - \frac{2a^2+bx}{\sqrt{a^2+bx}}$. We have $G'(b) = -\frac{b}{4(a^2+bx)^{3/2}} \leq 0$, implying $G$ is decreasing in $b \geq 0$. Thus, to show that $F'(x) \leq 0$, it suffices to show $\lim_{b\to 0} F'(x) \leq 0$, which trivially holds given $\lim_{b\to 0} F'(x) = \frac{1}{2x^2}(2a - \frac{2a^2}{a}) = \frac{1}{2x^2}(2a - 2a) = 0$. We can also verify $\lim_{b\to\infty} F'(x) = -\infty$, and thus $F'(x) \in (-\infty, 0]$ for $a, b, x \geq 0$. We thus have

$$f_i(\boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x\right)^{1/2} - \rho_q}{\frac{2}{\sqrt{\rho_i \tau_i}}x} \geq \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x'\right)^{1/2} - \rho_q}{\frac{2}{\sqrt{\rho_i \tau_i'}}x'} = f_i(\boldsymbol{\tau}', \boldsymbol{\rho}),$$

which completes the proof of part (i).

*Part (iii):* Suppose two players $i$ and $i'$ have the same precisions, $\rho_i = \rho_{i'}$, with $\tau_i > \tau_{i'}$. Then,

$$f_i(\boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau_i)^{1/2}}x} > \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau_{i'})^{1/2}}x} = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_{i'} \tau_{i'})^{1/2}}x} = f_{i'}(\boldsymbol{\tau}, \boldsymbol{\rho}).$$

*Part (iv):* Suppose two players $i$ and $i'$ have the same token holdings, $\tau_i = \tau_{i'}$, with $\rho_i > \rho_{i'}$. Then,

$$f_i(\boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i \tau_i)^{1/2}}x} > \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_i' \tau_i)^{1/2}}x} = \frac{\left(\rho_q^2 + \frac{4}{k^{1/2}}x\right)^{1/2} - \rho_q}{2\frac{1}{(\rho_{i'} \tau_{i'})^{1/2}}x} = f_{i'}(\boldsymbol{\tau}, \boldsymbol{\rho}).$$

*Part (ii):* To illustrate that effort is non-monotonic in precision, it suffices to consider the simplified case when $\rho_q \to 0$. In this case,

$$\frac{\partial f_i}{\partial \rho_i} = \frac{\tau_i \sqrt{\frac{x}{\sqrt{k}}}\left(x - \rho_i \frac{\partial x}{\partial \rho_i}\right)}{2\sqrt{\rho_i \tau_i}x^2},$$

where we now highlight that the sum $x$ depends on the information precisions $x(\rho_1, \ldots, \rho_n)$. The sign of $\frac{\partial f_i}{\partial \rho_i}$ is given by that of $x - \rho_i \frac{\partial x}{\partial \rho_i} \stackrel{\text{def}}{=} H(\rho_i)$. We have that $\frac{\partial H(\rho_i)}{\partial \rho_i} = -\rho_i \frac{\partial^2 x}{\partial \rho_i^2}$. Given $\frac{\partial^2 x}{\partial \rho_i^2} = \frac{3}{4}\frac{\rho_i \tau_i}{\sqrt{\rho_i^3 \tau_i}} \geq 0$, $\frac{\partial H(\rho_i)}{\partial \rho_i} \leq 0 \Rightarrow H(\rho_i)$ is decreasing, with $H(0) \geq 0$ (the inequality is strict whenever at least one other player has non-zero tokens and non-zero precision) and $\lim_{\rho_i \to \infty} H(\rho_i) = -\infty$. Solving $H(\rho_i) = 0$ gives the unique cutoff point where the derivative changes signs. □

**Proof of Proposition 6.** Part (i): The centralized platform's optimization problem is

$$\min_{\boldsymbol{u}\geq 0} V(\boldsymbol{u}) + \sum_{i=1}^{n} c(u_i).$$

The objective function is convex as long as $V'' > 0$ and $c'' > 0$, hence we search for a minimum through the FOC which are $V'(\boldsymbol{u}) + c'(u_i) = 0, \forall i$. By identification, comparing to the FOC of (10) after setting $\tau_i = 1, \forall i$, we obtain that under the assumptions of Theorem 1, the optimal solution of the centralized platform is $u_i^{fb}(\boldsymbol{\rho}) = u_i^*(\mathbf{1}, \boldsymbol{\rho})$, where $u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho})$ is given by (14).

Part (ii) follows from Proposition 5(i) that $u_i^*$ is increasing in $\tau_i$. By definition, $\tau_i \leq 1, \forall i$, hence $u_i^*(\boldsymbol{\tau}, \boldsymbol{\rho}) \leq u_i^*(\mathbf{1}, \boldsymbol{\rho}) = u_i^{fb}, \forall i$. If in addition $\tau_i \neq 1, \forall i$, the inequality is strict.

Part (iii) follows from (ii), and from (33), that $V' < 0$. □

**Proof of Proposition 7.** When all players have the same precision $\rho = 1/\sigma^2$, the equilibrium effort expression in (14), and (18) simplifies to

$$u_i^* = \frac{2\sigma_q^2}{\sigma\sqrt{\frac{k}{\tau_i}}\left(\sigma_q^2\sqrt{\frac{4(\sqrt{\tau_1}+\cdots+\sqrt{\tau_n})}{\sqrt{k}\sigma^3} + \frac{1}{\sigma_q^4}} + 1\right)} - 1.$$

We suppress the $*$ superscript to ease notation. Taking the limit $\sigma_q \to \infty$, and grouping terms, we obtain

$$\lim_{\sigma_q\to\infty} u_i = \sigma^{1/2}k^{-1/4}\tau_i^{1/2}S^{-1/2} - 1,$$

with $S = (\sqrt{\tau_1} + \cdots + \sqrt{\tau_n})$. It is useful to define the total effort exerted by all agents on the platform given by

$$u_{tot} = \sum_{i=1}^{n} u_i = \sigma^{1/2}k^{-1/4}S^{-1/2}\sum_{i=1}^{n}\tau_i^{1/2} - n = \sigma^{1/2}k^{-1/4}S^{-1/2}S - n = \sigma^{1/2}k^{-1/4}S^{1/2} - n.$$

The platform variance simplifies to

$$V(\boldsymbol{u}) = \frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{n} g(u_i)},$$

with $g(u) = 1 + u$ in the linear model (letting arbitrary constants $\eta_i = 1$). We have

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} g(u_i) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(1+u_i) = \frac{1}{\sigma^2}(n + u_{tot}) = \sigma^{-3/2}k^{-1/4}S^{1/2}.$$

Therefore, the platform variance can be written as

$$V = \sigma^{3/2}k^{1/4}S^{-1/2},$$

with $S = (\sqrt{\tau_1} + \cdots + \sqrt{\tau_n})$. Now, consider the problem of minimizing $V = V(\boldsymbol{\tau})$ over $\boldsymbol{\tau}$ with linear constraint $\sum \tau_i = 1$. Given the inverse relationship between $V$ and $S$, and the monotone transformation $\sqrt{(\cdot)}$, the problem is equivalent to maximizing $S(\boldsymbol{\tau})$ over $\boldsymbol{\tau}$ with a linear constraint. We have $\partial^2_{\tau_i} S = -\frac{1}{4}\tau_i^{-3/2}$ implying the Hessian of $S$ is negative definite, and therefore the problem is strictly concave. It is relatively straightforward to verify that the optimal solution is $\tau_i^* = 1/n$. By definition of concavity, any vector $\boldsymbol{\tau}$ that deviates from this optimal solution leads to lower $S$, that is, higher platform variance. □

### B.3 Proofs for Auxiliary Results in Appendix A

**Proof of Proposition 8.** A direct calculation shows that in this setting, $q - \hat{q}$ is normally distributed with mean 0, and variance

$$\mathrm{Var}\,(q - \hat{q}) = \left(1 - \sum_{i=1}^{n} w_i \alpha_i\right)^2 \sigma_q^2 + \sum_{i=1}^{n} \tau_i^2 \alpha_i^2 \sigma_i^2.$$

If the aggregator sets

$$w_i = \frac{\sigma_q^2}{\alpha_i \sigma_i^2 \left(1 + \sum_{j=1}^{n} \frac{\sigma_q^2}{\sigma_j^2}\right)},$$

where $\alpha_i = \beta_i$ for truthful voters, we have

$$\mathrm{Var}(q - \hat{q}) = \frac{1}{\sigma_q^{-2} + \sum_{i=1}^{n} \sigma_i^{-2}}.$$

On the other hand, by identification with (5), this is in fact the minimal possible variance given $s_1, \ldots, s_n$. Finally, Lemma 1 shows that if the payoff function is well-behaved, then choosing $w_i$ to maximize the payoff is equivalent to choosing $w_i$ to minimizing the variance $\mathrm{Var}(q - \hat{q})$. Since this strategy gives the minimum variance, it must also give the maximum possible payoff. □